METHOD FOR EXPRESSION OF HETEROLOGOUS PROTEINS IN YEAST

Inventor: Patricia Tekamp-Olson, San Anselmo, CA (US)

Assignee: Chiron Corporation, Emeryville, CA (US)

Notice: Subject to any disclaimer, the term of this patent is extended or adjusted under 35 U.S.C. 154(b) by 0 days.

This patent is subject to a terminal disclaimer.

Appl. No.: 09/921,398
Filed: Aug. 2, 2001

Prior Publication Data
US 2002/005169 A1 May 9, 2002

Related U.S. Application Data
Continuation of application No. 09/528,108, filed on Mar. 17, 2000, now Pat. No. 6,312,923, which is a continuation of application No. 09/540,250, filed on Jul. 1, 1999, now Pat. No. 6,083,723, which is a continuation of application No. 08/989,251, filed on Dec. 12, 1997, now Pat. No. 6,017,731.

Provisional application No. 60/032,720, filed on Dec. 13, 1996.

Field of Search
C12N 15/00
U.S. Cl. 435/69.4; 435/320; 435/254.2; 536/23.1; 536/24.1

References Cited
U.S. PATENT DOCUMENTS
4,769,328 A 9/1988 Murray et al.
4,845,075 A 7/1989 Murray et al.
4,889,919 A 12/1989 Murray et al.
5,010,003 A 4/1991 Chang et al.
5,219,739 A 6/1993 Heldin et al.
5,516,896 A 5/1996 Murray et al.
5,602,034 A 2/1997 Tekamp-Olson
6,083,723 A * 7/2000 Tekamp-Olson .......... 435/69.4
6,312,923 B1 * 11/2001 Tekamp-Olson .......... 435/69.1

FOREIGN PATENT DOCUMENTS
EP 0 206 783 A2 12/1986
EP 0 324 274 A1 7/1989
WO WO 95/02059 A1 1/1995

OTHER PUBLICATIONS

Primary Examiner—David Guzo
Assistant Examiner—Celine Qian
Attorney, Agent, or Firm—Robert P. Blackburn; Lisa E. Alexander; Leslie T. Henry

ABSTRACT
Compositions and methods for expression of heterologous mammalian proteins and their secretion in the biologically active mature form using a yeast host cell as the expression system are provided. Compositions of the invention are nucleotide sequences encoding a signal peptide sequence for a yeast secreted protein, an optional leader peptide sequence for a yeast secreted protein, a native propeptide leader sequence for a mature protein of interest, and a sequence for the mature protein of interest, all operably linked to a yeast promoter. Each of these elements is associated with a processing site recognized in vivo by a yeast proteolytic enzyme. Any or all of these processing sites may be a preferred processing site that has been modified or synthetically derived for more efficient cleavage in vivo. The compositions are useful in methods for expression of heterologous mammalian proteins and their secretion in the biologically active mature form. Particularly, vectors comprising these nucleotide coding sequences can be used to transform a yeast host cell, which can then be cultured and screened for secretion of the biologically active mature protein of interest.

15 Claims, 6 Drawing Sheets
OTHER PUBLICATIONS

* cited by examiner
Figure 1
Figure 2
Figure 3
Figure 4
Figure 5
METHOD FOR EXPRESSION OF HETEROLOGOUS PROTEINS IN YEAST

Cross-reference to related applications

The present application is a continuation of U.S. application Ser. No. 09/528,108, filed Mar. 17, 2000, now U.S. Pat. No. 6,312,923; which is a continuation of U.S. application Ser. No. 09/340,250, filed Jul. 1, 1999, now U.S. Pat. No. 6,083,723; which is a continuation of U.S. application Ser. No. 08/989,251, filed Dec. 12, 1997, now U.S. Pat. No. 6,017,731, which claims the benefit of U.S. Provisional Application Serial No. 60/032,720, filed Dec. 13, 1996, herein incorporated by reference in their entirety.

Field of the invention

The present invention relates to the production of recombinant proteins using yeast host cells as the expression system. More particularly, it relates to compositions and methods for expression of heterologous proteins and their secretion as the biologically active mature proteins.

Background of the invention

Yeast host expression systems have been used to express and secrete proteins foreign to yeast. Numerous approaches have been developed in terms of the degree of expression and the yield of biologically active mature proteins.

Such approaches have involved modifications to the various molecular components that are involved in expression and secretion of proteins in yeast. These components include the translation and termination regulatory regions for gene expression, signal peptide and secretion leader peptide sequences, which direct the precursor form of the heterologous protein through the yeast secretory pathway, and processing sites, which cleave leader peptide sequences from the polypeptide sequence of the protein of interest.

Expression of the protein of interest can be enhanced with use of yeast-recognized regulatory regions. Increased yield of the heterologous protein of interest is commonly achieved with the use of yeast-derived signal and secretion leader peptide sequences. The use of native signal-leader peptide sequences is believed to improve direction of the protein of interest through the secretory pathway of the yeast host.

Previous work has demonstrated that full-length yeast a-factor signal-leader sequences can be used to drive expression and processing of heterologous proteins in yeast host cells. Substantial improvements in efficiency of expression can be accomplished with the use of truncated a-factor leader sequences, particularly for heterologous proteins that are poorly expressed by the full-length sequence or whose expression is nonresponsive to the full-length sequence.

Although the various approaches available in the art have been shown to work with some proteins, problems persist with post-translational processing. Often the amount of protein secreted is unacceptably low or incorrect processing leads to inactive forms of the protein. This is particularly true for proteins that are initially expressed as a precursor polypeptide sequence and whose assumption of a native conformation is facilitated by the presence of a native propeptide sequence in the precursor polypeptide.

Methods for expression of heterologous proteins and their secretion in a biologically active mature form using a yeast host cell as the expression system are needed.

Summary of the invention

Compositions and methods for expression of heterologous proteins, more particularly heterologous mammalian proteins, and their secretion in a biologically active mature form using a yeast host cell as the expression system are provided. Compositions of the invention are nucleotide sequences encoding a signal sequence for a yeast secreted protein, a native propeptide leader sequence for a mature protein of interest, and a peptide sequence for the mature protein of interest. Each of these elements is associated with a processing site recognized in vivo by a yeast proteolytic enzyme. Any or all of these processing sites may be a preferred processing site that has been modified or synthetically derived for more efficient cleavage in vivo. In turn, all of these elements are operably linked to a yeast promoter and optionally other regulatory sequences.

The nucleotide coding sequences of these compositions may additionally comprise a leader peptide sequence for a yeast secreted protein. When present, this element, which is also associated with a processing site recognized in vivo by a yeast proteolytic enzyme, is positioned 3' to the yeast signal sequence and 5' to the sequence for the mature protein of interest. Thus cleavage by a yeast proteolytic enzyme removes the yeast leader sequence from the hybrid precursor molecule comprising the sequence for the mature protein of interest.

These compositions are useful in methods for expression of heterologous mammalian proteins and their secretion in the biologically active mature form. Particularly, vectors comprising these nucleotide coding sequences can be used to transform a yeast host cell, which can then be cultured and screened for secretion of the biologically active mature protein of interest.

The method of the present invention is particularly useful in production of mammalian proteins whose assumption of a native conformation is facilitated by the presence of a native propeptide sequence in the precursor polypeptide.

Detailed description of the drawings

FIG. 1 is a map of plasmid pAB24.
FIG. 2 is a map of the rhPDGF-B expression cassette in pAGL7PB and pYGGL7PB.
FIG. 3 is a map of rhPDGF-B expression plasmid pYGGL7PB.
FIG. 4 is a map of the rhPDGF-B expression cassette in pL7PPB and pYL7PPB.
FIG. 5 shows the final steps in the construction of the rhPDGF-B expression cassette in pL7PPB.
FIG. 6 is a map of rhPDGF-B expression plasmid pYL7PPB.

Detailed description of the invention

The present invention provides compositions and methods for expression of heterologous proteins of interest, more particularly heterologous mammalian proteins, and their secretion in a biologically active mature form using a yeast host cell as the expression system. By "biologically active mature form" is intended a protein whose conformational form is similar to the native conformation such that its biological activity is substantially the same as the biological activity of the native protein.

Compositions of the present invention are nucleotide sequences encoding hybrid precursor polypeptides that each comprise the polypeptide sequence for a mature heterologous protein of interest. Expression vectors comprising these nucleotide sequences, all under the operational control of a yeast promoter region and a yeast terminator region, are...
also provided. Methods of the invention comprise stably transforming a yeast host cell with said vectors, where expression of the nucleotide sequence encoding the hybrid precursor polypeptide leads to secretion of the mature heterologous protein of interest in a biologically active form.

By "heterologous protein of interest" is intended a protein that is not expressed by the yeast host cell in nature. Preferably the heterologous protein will be a mammalian protein, including substantially homologous and functionally equivalent variants thereof. By "variant" is intended a polypeptide derived from the native polypeptide by deletion (so-called truncation) or addition of one or more amino acids to the N-terminal and/or C-terminal end of the native protein; deletion or addition of one or more amino acids at one or more sites in the native polypeptide; or substitution of one or more amino acids at one or more sites in the native polypeptide. Such variants may result from, for example, genetic polymorphism or from human manipulation. Methods for such manipulations are generally known in the art.

For example, amino acid sequence variants of the polypeptide can be prepared by mutations in the cloned DNA sequence encoding the native polypeptide of interest. Methods for mutagenesis and nucleotide sequence alterations are well known in the art. See, for example, Walker and Gaastre, eds. (1983) Techniques in Molecular Biology (Macmillan Publishing Company, New York); Kunkel (1985) Proc. Natl Acad. Sci. USA 82:488-492; Kunkel et al. (1987) Methods Enzymol. 154:367-382; Sambrook et al. (1989) Molecular Cloning: A Laboratory Manual (Cold Spring Harbor Laboratory Press, Plainview, N.Y.); U.S. Pat. No. 4,873,192; and the references cited therein; herein incorporated by reference. Guidance as to appropriate amino acid substitutions that do not affect biological activity of the protein of interest may be found in the model of Dayhoff et al. (1978) in Atlas of Protein Sequence and Structure (Nat. Biomed. Res. Found., Washington, D.C.), herein incorporated by reference. Conservative substitutions, such as exchanging one amino acid with another having similar properties, may be preferred. Examples of conservative substitutions include, but are not limited to,

Gly, Ala, Val, Thr, Pro, Glu, Asp, Ser, Gln, Gly, and Phe, Tyr, and Trp.

In constructing variants of the protein of interest, modifications will be made such that variants continue to possess the desired activity. Obviously, any mutations made in the DNA encoding the variant protein must not place the sequence out of reading frame and preferably will not create complementary regions that could produce secondary mRNA structure. See EP Patent Application Publication No. 75,444.

Thus proteins of the invention include the naturally occurring forms as well as variants thereof. These variants will be substantially homologous and functionally equivalent to the native protein. A variant of a native protein is "substantially homologous" to the native protein when at least about 80%, more preferably at least about 90%, and most preferably at least about 95% of its amino acid sequence is identical to the amino acid sequence of the native protein. A variant may differ by as few as 1, 2, 3, or 4 amino acids. By "functionally equivalent" is intended that the sequence of the variant defines a chain that produces a polypeptide having substantially the same biological activity as the native protein of interest. Such functionally equivalent variants that comprise substantial sequence variations are also encompassed by the invention. Thus a functionally equivalent variant of the native protein will have a sufficient biological activity to be therapeutically useful. By "therapeutically useful" is intended effective in achieving a therapeutic goal, as, for example, healing a wound.

Methods are available in the art for determining functional equivalence. Biological activity can be measured using assays specifically designed for measuring activity of the native protein, including assays described in the present invention. Additionally, antibodies raised against the biologically active native protein can be tested for their ability to bind to the functionally equivalent variant, where effective binding is indicative of a protein having a conformation similar to that of the native protein.

The nucleotide sequences encoding the mature heterologous proteins of interest can be sequences cloned from non-yeast organisms, or they may be synthetic variants derived sequences, usually prepared using yeast-preferred codons. Examples of heterologous proteins suitable for the invention include, but are not limited to transforming growth factor (TGF-alpha and TGF-beta), somatostatin (as in SRIF 1), parathyroid hormone, and more particularly platelet-derived growth factor (PDGF) and insulin growth factor (IGF), all of which have a native prosequence as part of the precursor protein.

Thus compositions of the present invention are nucleotide sequences encoding hybrid precursor polypeptides that each comprise the polypeptide sequence for a mature heterologous protein of interest or any substantially homologous and functionally equivalent variants thereof. More particularly, nucleotide sequences of the present invention encode in the 5' to 3' direction a hybrid precursor polypeptide comprising the following primary elements:

\[5':\text{SP-}\text{(PS)}_{n-1}\text{-(LP-PS)}_{n-2}\text{-(NPRO}_{\text{ATVAR}}\text{PS)}_{n-3}\text{-MHP-}\text{CPRO}_{\text{ATVAR}}\text{MP}_{\text{PS}}\text{-}\]

wherein:

SP comprises a signal peptide sequence for a yeast secreted protein;

PS comprises a processing site cleaved in vivo by a yeast proteolytic enzyme;

LP comprises a leader peptide sequence for a yeast secreted protein;

NPRO_{\text{ATVAR}} comprises a native N-terminal propeptide sequence of a mature heterologous protein of interest;

MHP comprises a peptide sequence for said mature heterologous protein of interest;

CPRO_{\text{ATVAR}} comprises a native C-terminal propeptide sequence of said mature heterologous protein of interest; and

n-1, n-2, n-3, and n-4 independently 0 or 1;

wherein said processing sites allow for proteolytic processing of said precursor polypeptide to said mature protein in vivo by a yeast host cell, and wherein at least n-3 or n-4=1.

As is the case for the heterologous protein of interest, each of the other elements present in the hybrid precursor polypeptide can be a known naturally occurring polypeptide sequence or can be synthetically derived, including any variants thereof that do not adversely affect the function of the element as described herein. By "adversely affect" is intended inclusion of the variant form of the element results in decreased yield of the secreted mature heterologous protein of interest relative to the hybrid precursor polypeptide comprising the native form of the element.

In constructing the nucleotide sequence encoding the hybrid precursor polypeptide, it is within skill in the art to employ adapters or linkers to join the nucleotide fragments
encoding the various elements of the precursor polypeptide. See, for example, Sambrook et al. (1989) Molecular Cloning: A Laboratory Manual (Cold Spring Harbor Laboratory Press, Plainview, N.Y.). Thus, the hybrid precursor polypeptide may comprise additional elements positioned 5′ or 3′ to any of the primary elements listed above, including the yeast leader peptide sequence and its associated yeast-recognized processing site when present.

For purposes of the present invention, SP is a prosequence that is an N-terminal sequence for the precursor polypeptide of the mature form of a yeast secreted protein. When the nucleic acid sequence encoding the hybrid precursor polypeptide is expressed in a transformed yeast host cell, the signal peptide sequence functions to direct the hybrid precursor polypeptide comprising the mature heterologous protein of interest into the endoplasmic reticulum (ER). Movement into the lumen of the ER represents the initial step into the secretory pathway of the yeast host cell. Although the signal peptide of the invention can be heterologous to the yeast host cell, more preferably the signal peptide will be native to the host cell.

The signal peptide sequence of the invention may be a known naturally occurring signal sequence or any variant thereof as described above that does not adversely affect the function of the signal peptide. Examples of signal peptides appropriate for the present invention include, but are not limited to, the signal peptide sequences for α-factor (see, for example, U.S. Pat. No. 5,602,034; Brake et al. (1984) Proc. Natl. Acad. Sci. USA 81:4642–4646; invertase (WO 84/01153); PHO5 (DK 3614/83); YAP3 (yeast aspartic protease 3; PCT Publication No. 95/02059); and BAR1 (PCT Publication No. 87/02670). Alternatively, the signal peptide sequence may be derived from genomic or cDNA libraries using hybridization probe techniques available in the art (see Sambrook et al. (1989) Molecular Cloning: A Laboratory Manual (Cold Spring Harbor Laboratory Press, Plainview, N.Y.), or even synthetically derived (see, for example, WO 92/11378).

During entry into the ER, the signal peptide is cleaved off the precursor polypeptide at a processing site. The processing site can comprise any peptide sequence that is recognized in vivo by a yeast proteolytic enzyme. This processing site may be the naturally occurring processing site for the signal peptide. More preferably, the naturally occurring processing site will be modified, or the processing site will be synthetically derived, so as to be a preferred processing site. By “preferred processing site” is intended a processing site that is cleaved in vivo by a yeast proteolytic enzyme more efficiently than is the naturally occurring site. Examples of preferred processing sites include, but are not limited to, dibasic peptides, particularly any combination of the two basic residues Lys and Arg, that is Lys-Lys, Lys-Arg, Arg-Lys, or Arg-Arg, most preferably Lys-Arg. These sites are cleaved by the endopeptidase encoded by the KEX2 gene of Saccharomyces cerevisiae (see Fuller et al. Microbiology 1986:273–278) or the equivalent protease of other yeast species (see Julius et al. (1983) Cell 32:839–852). In the event that the KEX2 endopeptidase would cleave a site within the polypeptide sequence for the mature heterologous protein of interest, other preferred processing sites could be utilized such that the polypeptide sequence of interest remains intact (see, for example, Sambrook et al. (1989) Molecular Cloning: A Laboratory Manual (Cold Spring Harbor Laboratory Press, Plainview, N.Y.).

A functional signal peptide sequence is essential to bring about extracellular secretion of a heterologous protein from a yeast cell. Additionally, the hybrid precursor polypeptide may comprise a secretion leader peptide sequence of a yeast secreted protein to further facilitate this secretion process. When present, the leader peptide sequence is generally positioned immediately 3′ to the signal peptide sequence processing site. By “secretion leader peptide sequence” (LP) is intended a peptide that directs movement of a precursor polypeptide, which for the purposes of this invention is the hybrid precursor polypeptide comprising the mature heterologous protein to be secreted, from the ER to the Golgi apparatus and from there to a secretory vesicle for secretion across the cell membrane into the cell wall area and/or the growth medium. The leader peptide sequence may be native or heterologous to the yeast host cell but more preferably is native to the host cell.

The leader peptide sequence of the present invention may be a naturally occurring sequence for the same yeast secreted protein that served as the source of the signal peptide sequence, a naturally occurring sequence for a different yeast secreted protein, or a synthetic sequence (see, for example, WO 92/11378), or any variants thereof that do not adversely affect the function of the leader peptide.

For purposes of the invention, the leader peptide sequence when present is preferably derived from the same yeast secreted protein that served as the source of the signal peptide sequence, more preferably an α-factor protein. A number of genes encoding precursor α-factor proteins have been cloned and their combined signal-leader peptide sequences identified. See, for example, Singh et al. (1983) Nucleic Acids Res. 11:4049–4063; Kurjan et al., U.S. Pat. Nos. 4,546,082; 5,010,182; herein incorporated by reference. Alpha-factor signal-leader peptide sequences have been used to express heterologous proteins in yeast. See, for example, Elliott et al. (1983) Proc. Natl. Acad. Sci. USA 80:7080–7084; Bitter et al. (1984) Proc. Natl. Acad. Sci. 81:5330–5334; Smith et al. (1985) Science 229:1219–1229; and U.S. Pat. Nos. 4,849,407 and 5,219,759; herein incorporated by reference.

Alpha-factor, an oligopeptide mating pheromone approximately 13 residues in length, is produced from a larger precursor polypeptide of between 100 and 200 residues in length, more typically about 120–160 residues. This precursor polypeptide comprises the signal sequence, which is about 19–23 (more typically 20–22 residues), the leader sequence, which is about 60 residues, and typically 2–6 tandem repeats of the yeast α-factor sequence. Although the signal peptide sequence and full-length α-factor leader peptide sequence can be used, more preferably for this invention a truncated α-factor leader peptide sequence will be used with the signal peptide when both elements are present in the hybrid precursor molecule.

By “truncated” α-factor leader peptide sequence is intended a portion of the full-length α-factor leader peptide sequence that is about 20 to about 60 amino acid residues, preferably about 25 to about 50 residues, more preferably about 30 to about 40 residues in length. Methods for using truncated α-factor leader sequences to direct secretion of heterologous proteins in yeast are known in the art. See particularly U.S. Pat. No. 5,602,034. When the hybrid precursor polypeptide sequence comprises a truncated α-factor leader peptide, deletions to the full-length leader will preferably be from the C-terminal end and will be done in such a way as to retain at least one glycosylation site (-Asn-Y-Thr/Ser-, where Y is any amino acid residue) in the truncated peptide sequence. This glycosylation site, whose modification is within skill in the art, is retained to facilitate secretion (see particularly WO 89/02463).

When the hybrid precursor polypeptide sequence of the present invention comprises a leader peptide sequence, such
as the α-factor leader sequence, there will be a processing site immediately adjacent to the 3' end of the leader peptide sequence. This processing site enables a proteolytic enzyme native to the yeast host cell to cleave the yeast secretion leader peptide sequence from the 5' end of the native N-terminal propeptide sequence of the mature heterologous protein of interest, when present, or from the 5' end of the peptide sequence for the mature heterologous protein of interest. The processing site can be comprised by any peptide sequence that is recognized in vivo by a yeast proteolytic enzyme such that the mature heterologous protein of interest can be processed correctly. The peptide sequence for this processing site may be a naturally occurring peptide sequence for the native processing site of the leader peptide sequence. More preferably, the naturally occurring processing site will be modified, or the processing site will be synthetically derived, so as to be a preferred processing site as described above.

In the present invention, the nucleotide sequence encoding the hybrid precursor polypeptide comprises a native propeptide sequence (PRO<sub>α-factor</sub>) for the mature heterologous protein of interest. By “native propeptide sequence” or “native sequence for the mature heterologous protein of interest” is intended that portion of an intermediate precursor polypeptide (which is called a “proprotein”) for a mature secreted protein that remains attached to the N-terminal and/or C-terminal end of the mature protein sequence following cleavage of the native signal peptide sequence (or prosequence) from the initial proprotein polypeptide (or “proprotein”). The residues of the propeptide sequence are not contained in the mature secreted protein. Rather, such extra residues are removed at processing sites by proteolytic enzymes near the end of the secretory pathway, in the trans-Golgi network (Griffiths and Simons (1986) Science 234:438–443) and secretory granules (Orci et al. (1986) J. Cell Biol. 103:2273–2281).

The present invention provides for the presence of propeptide sequences that naturally occur at the N-terminal and/or C-terminal end of the native pro-protein precursor form of the mature heterologous protein of interest. Thus, a propeptide sequence may be positioned between the 3' end of the signal peptide sequence processing site, or the 3' end of the yeast-recognized processing site adjacent to the leader peptide sequence if present, and the 5' end of the peptide sequence for the mature heterologous protein of interest (a N-terminal propeptide sequence, PROM<sub>H</sub>) or immediately adjacent to the 3' end of the peptide sequence for the mature heterologous protein of interest (a C-terminal propeptide sequence, PRO<sub>α-factor</sub>), depending on its orientation within the native pro-protein. The invention also provides for inclusion of both an N-terminal and a C-terminal propeptide sequence flanking the peptide sequence for the mature heterologous protein of interest when both propeptide sequences exist in the native pro-protein. Where both an N-terminal and a C-terminal propeptide sequence exists in the native pro-protein, preference for inclusion of both propeptide sequences in the hybrid precursor polypeptide will be experimentally determined.


Immediately 3' to the native N-terminal propeptide sequence (when present) or immediately 5' to the C-terminal propeptide sequence (when present) is a processing site that is recognized in vivo by a yeast proteolytic enzyme. This processing site allows for cleavage of the propeptide sequence from the peptide sequence for the mature heterologous protein of interest (MHP). It is recognized that this processing site may be the naturally occurring processing site for the propeptide sequence if the naturally occurring site is recognized in vivo by a proteolytic enzyme of the yeast host cell. More preferably, the naturally occurring processing site will be modified, or the processing site will be synthetically derived, so as to be a preferred processing site. Examples of preferred processing sites include, but are not limited to, those discussed above for the other processing.

Preferably all of these processing sites will be similar such that the same yeast proteolytic enzyme brings about cleavage of the signal and leader peptide sequences and the native propeptide sequence(s).

In accordance with the invention as stated above, the yeast signal peptide and secretion leader peptide sequences, as well as the native propeptide sequences, represent those parts of the hybrid precursor polypeptide of the invention that can direct the sequence for the mature heterologous protein in vivo through the secretory pathway of a yeast host cell.

In one preferred embodiment of the present invention, the nucleotide sequence of the hybrid precursor polypeptide comprises in the 5' to 3' direction:

\[ 5'-\text{ASFSP-tAFLP-PS}_{\alpha}\text{-NPRO}_{\text{PDGF}}-\text{PS}_{\text{PDGF}}-\text{M}_{\text{PDGF}}-3' \]

wherein:

- ASFSP comprises an α-factor signal peptide sequence and a processing site;
- tAFLP comprises a truncated α-factor secretion leader peptide sequence;
- PS<sub>α</sub> comprises a preferred processing site for the leader peptide sequence;
- NPRO<sub>PDGF</sub> comprises the peptide sequence for a native N-terminal propeptide of a mature platelet-derived growth factor (PDGF);
- PS<sub>PDGF</sub> comprises a preferred processing site for the N-terminal propeptide sequence; and
- M<sub>PDGF</sub> comprises the sequence for said mature PDGF.

Preferably the α-factor signal peptide and truncated α-factor secretion leader peptide sequences are derived from the Mata gene of <i>S. cerevisiae</i> as outlined in the examples. The preferred truncated α-factor leader peptide sequence will include the N-terminal portion of the full-length leader sequence; that is, the leader sequence will start with the first amino acid residue of the full-length sequence and run the length of about 20 to about 60 amino acid residues, preferably about 25 to about 50 residues, more preferably about 30 to about 40 residues. In one embodiment, a leader of about 35 residues is used.

The mature protein of this preferred embodiment is human platelet-derived growth factor (PDGF). PDGF, the primary mitogen in serum for mesenchymal-derived cells, is stored in platelet alpha-granules. Injury to blood vessels activates the release of PDGF from these granules in the vicinity of the injured vessels. This mitogen acts as a potent chemotactic agent for fibroblasts and smooth muscle cells, as well as monocytes and neutrophils. The mitogenic activity of the localized PDGF results in proliferation of those cells at the site of injury, contributing to the process of wound repair.

Purified native platelet-derived growth factor (PDGF), a glycoprotein of about 30,000 daltons, is composed of two
disulfide-linked polypeptide chains. Two forms of these chains, designated A and B, have been identified. The native protein occurs as the homodimer AA or BB or the heterodimer AB, or a mixture thereof. A partial amino acid sequence for the PDGF-A chain has been identified (Johnsson et al. (1984) EMBO J. 3:921–928) and cDNAs encoding two forms of PDGF A-chain precursors have been described (U.S. Pat. No. 5,219,759). The A-chain is derived by proteolytic processing of a 211 amino acid precursor polypeptide. The cDNA encoding the PDGF-B chain has also been described (Nature (1985) 316:748–750). The B-chain is derived by proteolytic processing of a 241 amino acid precursor.

The mature PDGF protein of the present invention will be the biologically active dimeric form, including the homodimers PDGF-AA and PDGF-AB or the heterodimer PDGF-AB, and any substantially homologous and functionally equivalent variants thereof as defined above. For example, the native amino acid sequence for the A-chain or the B-chain may be truncated at either the N-terminal or C-terminal end. Thus removal of up to 15 or up to 10 amino acids from the N-terminal or C-terminal end, respectively, of the B-chain does not affect biological activity of the variant. Additionally, amino-acid substitutions may be made. For example, an amino acid such as serine may be substituted for any of the cysteine residues at positions 43, 52, 53, and 97 of the native human B-chain and at corresponding positions in the native A-chain to obtain substantially homologous and functionally equivalent variants of the native chain. Variants of the A-chain are known based on cloned DNA sequences, such as, for example, variants having an additional 6 or 19 amino acids at the C-terminal end. See, for example, Tong et al. (1987) Nature 328:619–621; Belscholz et al. (1986) Nature 320:695–699. One PDGF B-chain variant may be the corresponding substantially homologous portion of the amino-acid sequence encoded by the v-sis gene of simian sarcoma virus. The homologous region of the product of this gene, p28°, begins at amino acid 67 and continues to amino acid 175, and differs from the human B-chain by only 4 amino acid residues (see, for example, European Patent Application No. 0 487 116 A1). Functionally equivalent variants can be determined with assays for biological activity as described in the examples.

The nucleotide sequence encoding the mature PDGF protein of the present invention may be genomic, cDNA, or synthetic DNA. The genes encoding the native forms of PDGF have been sequenced, and several variants are well known in the art. Expression of PDGF homodimers and heterodimers is described in, for example, U.S. Pat. Nos. 4,766,073; 4,769,328; 4,801,542; 4,845,075; 4,849,407; 5,045,633; 5,128,321; and 5,187,263; herein incorporated by reference. Based on the known amino acid sequences for the A- and B-chain polypeptides, synthetic nucleotide sequences encoding PDGF A-chain and B-chain polypeptides may be made in vitro using methods available in the art. See particularly Sambrook et al. (1989) Molecular Cloning: A Laboratory Manual (Cold Spring Harbor Laboratory Press, Plainview, N.Y.). Where the mature protein of interest is the heterodimer PDGF-AB, the nucleotide sequences encoding the hybrid precursor polypeptides comprising the A- and B-chain polypeptides may be assembled as part of one expression cassette or assembled into separate expression cassettes for cotransformation of a yeast host cell.

In this preferred embodiment comprising mature PDGF, the C-terminal end of the truncated α-factor secretion leader peptide sequence and of the native N-terminal propeptide sequence will terminate in a preferred processing site, preferably a dibasic processing site that is specific for the KEX2 endoprotease of S. cerevisiae. The dipetides can be any combination of the basic residues Lys and Arg, more preferably a Lys-Arg dipetide.

The native prepro-PDGF-B additionally comprises a 51 amino acid C-terminal propeptide. In another preferred embodiment, the nucleotide sequence encoding the hybrid precursor polypeptide comprises in the 5' to 3' direction the following modified sequence:

5'-AFSP-ALFLP-SF5-PROPDFGSFSPFCSNSG-MPSG-SPSFCPSG-PGPOF-DOP-3'

wherein:

CPROPDFG comprises a C-terminal propeptide sequence for said PDGF mature heterologous protein of interest; and

PSCPRO comprises a preferred processing site for the C-terminal propeptide sequence.

Preferably the preferred processing site for the C-terminal propeptide sequence is similar to that of the leader peptide sequence and the N-terminal propeptide sequence, such that the same yeast proteolytic enzyme brings about cleavage of the α-factor leader peptide sequence and the sequences for both of the native propeptides. Inclusion of these two additional components is experimentally determined.

In another preferred embodiment of the invention, the nucleotide sequence of the hybrid precursor polypeptide comprises in the 5' to 3' direction:

5'-AFSP-ALFLP-SF5-MEAGPSFCPSG-PGPOF-DOP-3'

wherein:

AFSP comprises an α-factor signal peptide sequence and a processing site; AELF comprises an α-factor secretion leader peptide sequence; PSβ comprises a preferred processing site for the leader peptide sequence; MEG comprises the peptide sequence for a mature insulin-like growth factor (IGF); PSACPSP comprises a preferred processing site for the C-terminal propeptide sequence; and CPROIGF comprises the peptide sequence for a native C-terminal propeptide of said mature IGF.

Preferably the α-factor signal peptide and α-factor secretion leader peptide sequences are derived from the Matα gene of S. cerevisiae as outlined for the preferred embodiment for PDGF.

The mature protein of this preferred embodiment is insulin-like growth factor (IGF), more particularly IGF-I. Insulin-like growth factor (IGF-I) belongs to a family of polypeptides known as somatomedins. IGF-I stimulates growth and division of a variety of cell types, particularly during development. See, for example, European Patent Application Nos. 560,723 A and 436,469 B. Thus, processes such as skeletal growth and cell replication are affected by IGF-I levels.

IGF-I is structurally and functionally similar to, but antigenically distinct from, insulin. In this regard, IGF-I is a single-chain polypeptide with three intrachain disulfide bridges and four domains known as the A, B, C, and D domains, respectively. The A and B domains are connected by the C domain, and are homologous to the corresponding domains of proinsulin. The D domain, a C-terminal prosequence, is present in IGF-I but is absent from proins-

The mature IGF protein of the present invention will be the biologically active form and any substantially homologous and functionally equivalent variants thereof as defined above. Functionally equivalent variants can be determined with assays for biological activity, including the assay, as described in the examples. Representative assays include known radioreceptor assays using placental membranes (see, for example, U.S. Pat. No. 5,324,630; Hall et al. (1975) J. Clin. Endocrinol. and Metab. 39:973–976; and Marshall et al. (1974) J. Clin. Endocrinol. and Metab. 39:283–292), a bioassay that measures the ability of the molecule to enhance incorporation of tritiated thymidine, in a dose-dependent manner, into the DNA of BALB/c 3T3 fibroblasts (see, for example, Tamura et al. (1989) J. Biol. Chem. 262:5616–5621), and the like; herein incorporated by reference.


The nucleotide sequence encoding the mature IGF protein of the present invention may be genomic, cDNA, or synthetic DNA. The genes encoding the native forms of IGF-I have been sequenced, and several variants are well known in the art. IGF-I and variants thereof can be produced in any number of ways that are well known in the art. For example, IGF-I polypeptides can be isolated directly from blood, such as from serum or plasma, by known methods. See, for example, U.S. Pat. No. 4,760,361; Svatoboda et al. (1980) Biochemistry 19:790–797; and Boughladj (1982) Prep. Biochem. 12:57 and (1984) Prep. Biochem. 14:123; herein incorporated by reference. Alternatively, IGF-I can be synthesized chemically, by any of several techniques that are known to those skilled in the art. See, for example, Stewart and Young (1984) Solid Phase Peptide Synthesis (Pierce Chemical Company, Rockford, Ill.) and Barany and Merrifield (1980) The Peptides: Analysis, Synthesis, Biology (eds. Gross and Meienhofer) pp. 3–254, Vol. 2 (Academic Press, New York) for solid phase peptide synthesis techniques and Bodansky (1984) Principles of Peptide Synthesis (Springer-Verlag, Berlin) and Gross and Meienhofer, eds. (1980) The Peptides: Analysis, Synthesis, Biology, Vol. 1, for classical solution synthesis; herein incorporated by reference. The IGF-I polypeptides of the present invention can also be chemically prepared by the method of simultaneous multiple peptide synthesis. See, for example, Houghten (1985) Proc. Natl. Acad. Sci. USA 82:5131–5135; U.S. Pat. No. 4,631,211; herein incorporated by reference.

In this preferred embodiment comprising mature IGF-I, the C-terminal end of the truncated c-factor secretion leader peptide sequence and the N-terminal end of the native C-terminal propeptide sequence will terminate in a preferred processing site, preferably a dibasic processing site that is specific for the KEX2 endopeptidase of S. cerevisiae. The dipeptidylases can be any combination of the basic residues Lys and Arg, more preferably a Lys-Arg dipeptide.

The nucleotide sequences of the present invention are useful for producing biologically active mature heterologous proteins of interest in a yeast host cell when operably linked to a yeast promoter. In this manner, the nucleotide sequences encoding the hybrid precursor polypeptide of the invention are provided in expression cassettes for introduction into a yeast host cell. These expression cassettes will comprise a transcriptional initiation region linked to the nucleotide sequence encoding the mature polypeptide of interest, and a yeast expression cassette is provided with a plurality of restriction sites for insertion of the nucleotide sequence to be under the transcriptional regulation of the regulatory regions. The expression cassette may additionally contain selectable marker genes. Such an expression cassette comprises in the 5' to 3' direction and operably linked a yeast-recognized transcription and translation initiation region, a nucleotide coding sequence for the hybrid precursor polypeptide comprising the sequence for the mature protein of interest, and a yeast-recognized transcription and translation termination region. By “operably linked” is intended expression of the coding sequence for the hybrid precursor polypeptide is under the regulatory control of the yeast-recognized transcription and translation initiation and termination regions.

By “yeast-recognized transcription and translation initiation and termination regions” is intended regulatory regions that flank a coding sequence, in this case the nucleotide sequence encoding the hybrid polypeptide sequence, and control transcription and translation of the coding sequence in a yeast. These regulatory regions must be functional in the yeast host. The transcription initiation region, the yeast promoter, provides a binding site for RNA polymerase to initiate downstream (3') translation of the coding sequence. The promoter may be a constitutive or inducible promoter, and may be native or analogous or foreign or heterologous to the specific yeast host. Additionally, the promoter may be the natural sequence or alternatively a synthetic sequence. By foreign is intended that the transcription initiation region is not found in the native yeast of interest into which the transcription initiation region is introduced. Suitable native yeast promoters include, but are not limited to the wild-type c-factor promoter, as well as other yeast promoters. Preferably the promoter is selected from the list including promoters for the glycolytic enzymes phosphoglucoisomerase, phosphofructokinase, phosphotrioseisomerase, phosphoglucomutase, enolase, pyruvate kinase (PyK), glyceraldehyde-3-phosphate dehydrogenase (GAP or GAPDH), alcohol dehydrogenase (ADH) (EPO Publication No. 284,044). See, for example, EPO Publication Nos. 255,511 and 164,506.

Synthetic hybrid promoters consisting of the upstream activator sequence of one yeast promoter, which allows for inducible expression, and the transcription activation region
of another yeast promoter also serve as functional promoters in a yeast host. Examples of hybrid promoters include ADH/GAP, where the inducible region of the ADH promoter is combined with the activation region of the GAP promoter (U.S. Pat. Nos. 4,876,197 and 4,880,734). Other hybrid promoters using upstream activation sequences of either the ADH2, GAL4, GAL10, or PHO5 genes combined with the transcriptional activation region of a glycolytic enzyme such as GAP or Pyk are available in the art (EPO Publication No. 164,556). More preferably the yeast promoter is the inducible ADH/GAP hybrid promoter.


The termination regulatory region of the expression cassette may be native with the transcription initiation region, or nucleotide sequences may be synthesized using yeast sequences that is recognized by the yeast host. The termination regions may be those of the native α-factor transcription termination sequence, or another yeast-recognized termination sequence, as those for the glycolytic enzymes mentioned above. More preferably the transcription terminator is the Mata (α-factor) transcription terminator described in U.S. Patent 4,870,008.

The nucleotide sequences encoding the hybrid prucisor polypeptides of the invention are provided in expression cassettes for expression in a yeast host. The cassette will include 5' and 3' regulatory sequences operably linked to the nucleotide sequence encoding the hybrid prucisor polypeptide of interest. The cassette may also contain at least one additional nucleotide sequence of interest to be cotransformed into the yeast host. Alternatively, the additional nucleotide sequences can be provided on another expression cassette. Where appropriate, the nucleotide sequence encoding the hybrid prucisor polypeptide and any additional nucleotide sequences of interest may be optimized for increased expression in the transformed yeast. That is, these nucleotide sequences may be synthesized using yeast preferred codons for improved expression. Methods are available in the art for synthesizing yeast-preferred nucleotide sequences of interest (see, for example, U.S. Pat. Nos. 5,219,759 and 5,602,034).

Additional sequence modifications are known to enhance expression of nucleotide coding sequences in a cellular host. These include elimination of sequences encoding spurious polyadenylation signals, exon-intron splice site signals, transposon-like repeats, and other such well-characterized sequences that may be deleterious to gene expression. The G-C content of the sequence may be adjusted to levels of average for a given cellular host, as calculated by reference to known genes expressed in the host cell. When possible, the nucleotide coding sequence is modified to avoid predicted hairpin secondary mRNA structures.

In preparing the expression cassette, the various nucleotide sequence fragments may be manipulated, so as to provide for the sequences in the proper orientation and, as appropriate, in the proper reading frame. Toward this end, adaptors may be linked may be employed to join the nucleotide fragments or other manipulations may be involved to provide for convenient restriction sites, removal of superfluous nucleotides, removal of restriction sites, or the like. For this purpose, in vitro mutagenesis, primer repair, restriction, annealing, resubstitutions, e.g., transitions and transversions, may be involved. See particularly Sambrook et al. (1989) Molecular Cloning: A Laboratory Manual (Cold Spring Harbor Laboratory Press, Plainview, N.Y.).

The expression cassettes of the present invention can be ligated into a replicon (e.g., plasmid, cosmid, virus, minichromosome), thus forming an expression vector that is capable of autonomous DNA replication in vivo. Preferably the replicon will be a plasmid. Such a plasmid expression vector will be maintained in one or more replication systems, preferably two replication systems, that allow for stable maintenance within a yeast host cell for expression purposes, and within a prokaryotic host for cloning purposes. Examples of such yeast-bacteria shuttle vectors include YEp24 (Botstein et al. (1979) Gene 8:17–24; pC11 (Brake et al. (1984) Proc. Natl. Acad. Sci. USA 81:4642–4646), and Yrp17 (Stinchcomb et al. (1982) J. Mol. Biol. 158:157).

Additionally, a plasmid expression vector may be a high or low copy number plasmid, the copy number generally ranging from about 1 to about 200. With high copy number yeast vectors, the yeast vector is generally inserted at least 20, and usually not exceeding about 150 copies in a single host. Depending upon the heterologous protein selected, either a high or low copy number may be desirable, depending upon the effect of the vector and the foreign protein on the host. See, for example, Brake et al. (1984) Proc. Natl. Acad. Sci. USA 81:4642–4646. DNA constructs of the present invention can also be integrated into the yeast genome by an integrating vector. Examples of such vectors are known in the art. See, for example, Botstein et al. (1979) Gene 8:17–24.

The host chosen for expression of the heterologous proteins of the invention will preferably be a yeast. By “yeast” is intended ascosporesporogenous yeasts (Endomycetes), asidosporesporogenous yeasts, and yeast belonging to the Fungi Imperfecti (Blastomyces). The asidosporesporogenous yeasts are divided into two families, Spermophthoraeceae and Saccharomyceae. The later is comprised of four subclasses, Schizosaccharomyceae (e.g., genus Schizosaccharomyces), Nadsonioideae, Lipomyceae, and Saccharomyceae (e.g., genera Pichia, Kluyveromyces, and Saccharomyces). The basidiosporogenous yeasts include the genera Leucosporidium, Rhodosporidium, Sporidiobolus, Filobasidium, and Filobasidiella. Yeast belonging to the Fungi Imperfecti are divided into two families, Sporobolomyceae (e.g., genera Sporobolomyces, Bullera) and Cryptococcaceae (e.g., genus Candida). Of particular interest to the present invention are species within the genera Pichia, Kluyveromyces, Saccharomyces, Schizosaccharomyces, and Candida. Of particular interest are the Saccharomyces species S. cerevisiae, S. carlsbergensis, S. diastattes, S. douglasi, S. kluveri, S. norbensis, and S. oviformis. Species of particular interest in the genus Kluyveromyces include K. lactis. Since the classification of yeast may change in the future, for the purposes of this invention, yeast shall be defined as described in Skinner et al., eds. (1980) Biology and activities of Yeast (Soc. App. Bacteriol. Symp. Series No. 9). In addition to the foregoing, those of ordinary skill in the art are presumably familiar with the biology of yeast and the manipulation of yeast genetics. See, for example, Bacila et al., eds. (1978) Biochemistry and Genetics of Yeast; Rose and Harrison, eds. (1987) The Yeast (2nd ed.); Strathem et al., eds. (1981) The Molecular Biology of the Yeast Saccharomyces; herein incorporated by reference.
The selection of suitable yeast and other microorganism hosts for the practice of the present invention is within the skill of the art. When selecting yeast hosts for expression, suitable hosts may include those shown to have, inter alia, good secretion capacity, low proteolytic activity, and overall vigor. Yeast and other microorganisms are generally available from a variety of sources, including the Yeast Genetic Stock Center, Department of Biophysics and Medical Physics, University of California, Berkeley, Calif.; and the American Type Culture Collection, Rockville, Md.

Methods of introducing exogenous DNA into yeast hosts are well known in the art. There is a wide variety of ways to transform yeast. For example, spheroplast transformation is taught by Himen et al. (1978) Proc. Natl. Acad. Sci. USA 75:1919–1933 and Stinchcomb et al., EPO Publication No. 45,573; herein incorporated by reference. Transformants are grown in an appropriate nutrient medium, and, where appropriate, maintained under selective pressure to insure retention of endogenous DNA. Where expression is inducible, growth can be permitted of the yeast host to yield a high density of cells, and then expression is induced. The secreted, mature heterologous protein can be harvested by any conventional means, and purified by chromatography, electrophoresis, dialysis, solvent-solvent extraction, and the like.

The following examples are offered by way of illustration and not by way of limitation.

EXPERIMENTAL

The following examples further describe the construction of an expression vector comprising the nucleotide sequence encoding mature human PDGF-B in accordance with the disclosed invention. Examples demonstrating the use of this expression vector to produce biologically active mature PDGF-BB in a yeast host are also provided. Additional examples describe an expression vector comprising the nucleotide sequence encoding mature human IGF-I in accordance with the disclosed invention and demonstrate the use of this expression vector to produce biologically active mature IGF-I in a yeast host.

Example 1

Plasmid Vector pAB24

The vector selected for expressing rPDGF-BB, pAB24, is a yeast-bacteria shuttle vector. The plasmid is a chimera of sequences from pBR322, derived from several naturally occurring bacterial plasmids, and sequences of the endogenous S. cerevisiae 2-u plasmid (Brouoch (1981) in Molecular Biology of the Yeast Saccharomyces (Cold Spring Harbor Press, New York), 1:445–470). It also encodes genes enabling selection in both E. coli and S. cerevisiae hosts. The pBR322 part of pAB24 includes the ampicillin resistance (Ap')-confering gene encoding β-lactamase, as well as a gene conferring tetracycline resistance (Tc'). These genes allow transformation of competent E. coli and selection of plasmid-containing bacteria. A unique BamHI cloning site, present in the gene encoding tetracycline resistance, is the site utilized for insertion of an expression cassette. The pBR322 portion of the vector also includes a ColE1-like replication origin enabling replication in E. coli. Two S. cerevisiae genes derived from YEp24 (Boehn et al. (1979) Gene 8:17–24), URA3 and leu2d, enable selection in yeast host strains lacking either or both of these genes. The latter gene, leu2d, lacks a portion of the 5'-untranslated promoter region and requires high plasmid copy number for growth in leucine-deficient medium. This is necessary to achieve sufficient LEU2 protein expression for complementation of yeast strains lacking LEU2 (Ehrhart and Hollenberg (1983) J. Bacteriol. 156:625–635). The 2-μ sequences of pAB24 confer replication and partitioning of the expression plasmid in S. cerevisiae. FIG. 1 shows a schematic map of plasmid pAB24 with key restriction sites and genetic elements. A description of the construction of pAB24 can be found in the European Patent Application publication EPO 0324 274 B1.

Three expression plasmids containing the PDGF-B gene, pYAGL7PB, pYL7PB (also known as pYAGL7PB), and pYJST400, were used to produce PDGF-BB in a yeast host. All of these expression vectors utilize pAB24 as the plasmid into which the expression cassette comprising the PDGF-B gene was inserted.

Example 2

Construction of Expression Plasmid pYAGL7PB

General Description

Plasmid pYAGL7PB includes an expression cassette with the following features. Transcription is mediated by the inducible, hybrid yeast promoter ADH1/GAP. This promoter includes ADR2 transcription factor responsive sequences from the S. cerevisiae ADH2 gene (Beier and Young (1982) Nature 300:724–728) and promoter sequences from the S. cerevisiae gene TDH3, encoding the glycolytic enzyme glyceraldehyde-3-phosphate dehydrogenase (GAP). The ADR2 transcription factor responsive sequences confer inductive gene transcription upon downstream sequences. Induction is achieved by glucose depletion in the growth medium. Termination of transcription is mediated by the terminator derived from the S. cerevisiae mating factor type alpha (Mato) gene (Brake et al. (1984) Proc. Natl. Acad. Sci. USA 81:4642–4646).

The cassette further includes an open reading frame encoding a truncated Mata sequence fused to a sequence encoding the human PDGF-B gene. The truncated α-factor leader mediates secretion of in-frame protein fusions. It is a derivative of S. cerevisiae α-factor leader, the product of the Mata gene (Kurjan and Herskowitz (1982) Cell 30:933–943). A dibasic amino acid processing site is present at the truncated α-factor leader/PDGF-B junction to facilitate production of correctly processed rPDGF-BB polypeptide by yeast. FIG. 2 shows a map of the pYAGL7PB expression cassette highlighting these features and the restriction enzyme sites relevant to the construction of this expression cassette. The nucleotide sequence and predicted amino acid sequence of the open reading frame encoding the truncated α-factor leader/PDGF-B primary translation product are given in SEQ ID NO: 1 and SEQ ID NO: 2, respectively.

Sequential Construction of pYAGL7PB

Following is a description of the sequential steps taken to construct this expression vector.

Construction of PDGF-B Synthet Gene and Cloning into a Yeast Expression Vector

The synthetic gene encoding the partial dibasic processing site and rPDGF-B (SEQ ID NOs: 3–4) was made from 17 overlapping oligonucleotides (SEQ ID NOs: 5–21) as described in Udrea et. al. (Proc. Natl. Acad. Sci. USA 80 (1983):7461–7465). Ligation of the fragments results in an XbaI-Sall fragment, which was subsequently inserted into XbaI-Sall cut pYAG/αCF vector. Plasmid pYAG/αCF is a pBR322 derivative with an expression cassette delineated by BamHI sites. The expression cassette includes the ADH1/GAP hybrid promoter, as well as
the open reading frame encoding the yeast α-factor leader (BamHI-XbaI), an XbaI-SalI gene fragment, and the Matα (α-factor) transcription terminator (SalI-BamHI). Substitution of an XbaI-SalI gene fragment (in-frame) capable of heterologous protein expression into this plasmid allows the expression and secretion of the heterologous protein. The isolation of the yeast glycerolaldehyde-3-phosphate (GAP) gene promoter, the origin of the ADH2 component of the promoter, and the construction of a hybrid ADH1/GAP promoter are described in U.S. Pat. Nos. 4,876,197 and 4,880,734. The isolation of the yeast α-factor gene including the transcription terminator is described in U.S. Pat. No. 4,870,008.

Upon dideoxy sequencing, the synthetic gene sequence was found to have a single base pair mutation, which was repaired by standard procedures. Plasmid pPAGB8-B is the plasmid derived from pPAG/d/pd that contains the correct synthetic PDGF-B (XbaI-SalI) gene.

Construction of Synthetic Truncated α-Factor Leader Gene

With Dibasic Processing Site

The truncated α-factor leader mediates secretion of in-frame hybrid polypeptides. It is a derivative of S. cerevisiae α-factor leader, the secretion leader for mating factor type alpha, the product of the Matα gene (Kurjan and Herskowitz (1982) Cell 30:93:943), and consists of the first 35 amino acids of the native leader. The construction and use of a truncated α-factor leader gene to mediate secretion is described in EPO Publication No. 0324 274 B1. Synthetic oligonucleotides encoding a comparable, partial (amino acids 8–35) truncated α-factor leader (L7) and part of the dibasic processing site were made from oligonucleotides given in SEQ. ID NO: 22 and when assembled with the complementary strand shown in SEQ. ID NO: 23 resulted in a Psbl-BglII fragment with a 3′-ACGTG- and a 5′-CTAG-overhang to allow for convenient ligation into the expression cassette.

Construction of pAGL7PB

The purpose of this construction was the substitution of the synthetic, partial truncated α-factor leader Psbl-BglII gene fragment described above for most of the full-length α-factor leader in the PDGF-B expression cassette of pPAGB8-B. A 1.9 kb PstI fragment including prR322 sequences from the ADH-GAP hybrid promoter (marked at the 5′ end by a BamHI site) and the 5′ partial α-factor leader gene sequence (encoding the first seven amino acids of the native α-factor leader) was isolated from pPAGB8-B. It was ligated to kintaned, annealed synthetic oligonucleotides 1.49/3.4/40. Following digestion with BamHI, a partial expression cassette 5′ fragment was obtained including sequences for the ADH/GAP hybrid promoter and the 5′ portion of the truncated α-factor leader.

Similarly, a BglII fragment containing the PDGF-B synthetic gene, the α-factor terminator (marked at the 3′ end by a BamHI site) and prR322 sequences was isolated from pPAGB8-B. It was ligated to kintaned, annealed synthetic oligonucleotides 2.32/4.50. Following digestion with BamHI, a partial expression cassette 3′ fragment was obtained including sequences for the 3′ portion of the truncated α-factor leader, PDGF-B, and the α-factor leader transcription terminator. The complete PDGF-B expression cassette was obtained following ligation of the 5′ and 3′ partial expression cassette gene fragments and digestion with BamHI. The BamHI expression cassette was cloned into the BamHI site of a prR322-derived vector (pBR/ACEol-Sal) to give plasmid pAGL7PB. A map of the PDGF-B expression cassette in this plasmid is shown in FIG. 2.

Construction of pYAGL7PB

The PDGF-B expression cassette of pAGL7PB was isolated by BamHI digestion and inserted into the BamHI site of the yeast-bacteria shuttle vector pAB24 described above. A yeast expression plasmid, pYAGL7PB, was isolated. A plasmid map of pYAGL7PB is shown in FIG. 3. The nucleotide sequence of the complete expression cassette and the predicted amino acid sequence of the open reading frame (ORF) encoding the truncated α-factor leader-PDGF-B primary translation product are given in SEQ ID NO: 24 and SEQ ID NO: 25, respectively.

Expression Strain Identification: MB2-1pYAGL7PB

Expression plasmid pYAGL7PB was transformed into S. cerevisiae MB2-1 by standard procedures and prototrophic uracil colonies were selected. Individual colonies from independent transformants were screened for expression following inoculation of single colonies into medium that selects for leucine prototrophs. The medium also is high in glucose to keep expression of sequences under ADR2 regulation (including the PDGF-B gene) repressed. Cultures were subsequently diluted and grown to confluence in low glucose medium lacking uracil. Cell-free culture supernatants were prepared and assayed for PDGF-B by immunoactivity (ELISA) and by mitogenic activity on 3T3 cells. A high PDGF-B expressing colony, MB2-1(pYAGL7PB #5), was identified.

Example 3

Construction of Expression Plasmid pY77PPB

General Description

Plasmid pY77PPB (also known as pYAGL7PPB) includes an expression cassette with the following features. Transcription initiation and termination is mediated by the inducible, hybrid yeast promoter ADH1/GAP and the Matα transcriptional terminator described above. The gene further includes an open reading frame encoding a truncated yeast α-factor leader to mediate secretion of rhPDGF-BB. The propeptide sequence included in the expression construct is only the native N-terminal propeptide sequence; the native C-terminal propeptide sequence was not included in the construct. Inclusion of the N-terminal propeptide sequence resulted in enhanced expression of rhPDGF-BB, presumably because of improved folding. Dibasic processing sites at the truncated α-factor leader/N-terminal propeptide and N-terminal propeptide/PDGF-B junctions were included to facilitate production of correctly processed rhPDGF-BB polypeptide by yeast. FIG. 4 shows a map of the pY77PPB expression cassette highlighting these features and the sites relevant to the construction of this expression cassette. The nucleotide sequence and predicted amino acid sequence of the open reading frame encoding the truncated α-factor leader-proPDGF-B primary translation product are shown in SEQ ID NO: 26 and SEQ ID NO: 27, respectively.

Sequential Construction of pY77PPB

Source of rhPDGF-B-B cDNA

A cloned cDNA encoding native human preproPDGF-B, 8hPGFB-17, was provided by collaborators Arne Östman and Carl Heldin. Isolation of the cDNA encoding hPDGF-B was achieved using a cDNA library prepared from RNA isolated from a human clonal glioma cell line, U-343 MG. Cl 2. (Östman et al. (1988) J. Biol. Chem. 263:16202–16208). Construction of pSV7d-PDGF A103-B1

Plasmid pSV7d-PDGF A103-B1 was the source of the N-terminal propeptide-PDGF-B-B cDNA. The plasmid was constructed as described below.

The 3 kb Eco R1 PDGF-B-B cDNA insert from clone 8hPGFB-17 was excised and cloned into the unique Eco RI
site of the mammalian expression vector pSV7d to give plasmid pH hedgehog.

A mammalian plasmid, pSV7d-PDGF A103-B1, for the expression of both PDGF-A and -B chains from their respective cDNAs, was constructed as follows. Plasmid pH hedgehog was digested with PstI under conditions favoring cleavage at one of the two plasmid PstI sites (desired single cleavage at site in ampicillin resistance gene of the pSV7d vector backbone) and ligated with PstI-digested pSV7d-PDGF A103-B1 (D1). This latter plasmid is strictly analogous to the PDGF-B mammalian expression plasmid pH hedgehog, except that it includes cDNA encoding the long, 211 amino acid form of the PDGF-A chain rather than the PDGF-B chain cDNA. This plasmid contains a single PstI site in the ampicillin resistance gene of the pSV7d vector backbone.

Following transformation, bacterial colonies were screened for the presence of both PDGF-B and PDGF-A cDNA sequences with the respective or appropriately labeled EcoRI cDNA probes. Colonies positive for both PDGF-B and -A chain sequences were further screened by EcoRI digestion of plasmid DNA, and plasmid pSV7d-PDGF A103-B1, having a predicted EcoRI pattern, was identified.

**Mutagenesis of hPDGF-B cDNA**

The PDGF-B cDNA was mutagenized: (1) to introduce a SacI site enabling introduction of the truncated α-factor secretion leader, and (2) to change the hPDGF-B cDNA sequence encoding dibasic amino acids Arg-Arg to encode Lys-Arg. This dibasic combination is more efficiently cleaved than Arg-Arg by the yeast dibasic processing enzyme KEX2 endopeptidase. The template for mutagenesis was prepared as follows.

The ~3 kb EcoRI hPDGF-B cDNA was isolated from pSV7d-PDGF A103-B1 and inserted into the EcoRI site of pBR322 to give plasmid pPPB6. The nucleotide sequence of the 2.7 kb PstI-EcoRI cDNA fragment was verified. The 0.9 kb PstI-NcoI cDNA fragment was inserted into the PstI-NcoI site of M13 and the nucleotide sequence of the insert verified. A partial nucleotide sequence and the predicted amino acid sequence of the PDGF-B cDNA are given in SEQ ID NO: 28 and SEQ ID NO: 29, respectively.

A double mutagenesis of M13 PstI-NcoI PDGF-B cDNA fragment was performed by standard methods using the following primers: Primer 1 (SEQ ID NO: 30) introduces a SacI site; primer 2 (SEQ ID NO: 31) converts Arg-Arg to Lys-Arg at the propeptide/PDGF-B junction. Additional mutations are introduced to facilitate detection of mutagenized sequences by hybridization with the labeled primer. No changes resulted in the primary amino acid sequence by primer 1 mutagenesis; only the Arg=1 lysine amino acid change resulted from primer 2 mutagenesis. Mutant hPDGF-B inserts were detected by hybridization with both primer 1 and 2 radiolabeled probes. DNA sequence was verified, and RF (double-stranded) plasmid was prepared.

Construction of pL7PBP (pAGL7PBP)

Essentially, the steps described below result in the substitution of the XhoI-SalI portion of the PDGF-B expression cassette in pAGL7PBP encoding the C-terminal portion of the truncated α-factor leader, the Lys-Arg dibasic processing site and PDGF-B (FIG. 2) with an XhoI-SalI gene fragment encoding the C-terminal portion of the truncated α-factor leader, a Lys-Arg dibasic processing site, the PDGF-B N-terminal propeptide, a Lys-Arg dibasic processing site, and a PstI fragment. The sequences encoding the N-terminal PDGF-B propeptide and PDGF-B were derived from cDNA as described above. A map of the resulting expression cassette is shown in FIG. 4.

A 447 bp SacI-SphI fragment including most of the proPDGF-B gene was isolated from the M13 RF containing the modified preproPDGF-B cDNA. Synthetic oligonucleotides, including sequences encoding the C-terminal part of truncated α-factor leader, a Lys-Arg dibasic processing site, and the N-terminal portion of the PDGF-B propeptide (SEQ ID NO: 32–33), were joined to give a fragment with a 3′ SacI overhang. Synthetic oligonucleotides, XhoI-SalI (pAGL7PBP including sequences encoding the last 14 amino acids of PDGF-B and stop codons were ligated to give a SphI-SalI fragment (SEQ ID NO: 34–35). These two sets of annealed oligonucleotides were ligated to the 447 bp SacI-SphI proPDGF-B gene fragment. This resulted in a gene fragment including sequences encoding the C-terminal part of truncated α-factor leader, a Lys-Arg dibasic processing site, and proPDGF-B.

Synthetic oligonucleotides, including sequences encoding the middle amino acids of the treated α-factor leader were joined resulting in a fragment with a 5′ XhoI overhang (SEQ ID NO: 32–33). This annealed oligonucleotide was ligated with pAGL7PBP that had been cut with XhoI (unique site in pAGL7PBP plasmid that is in the expression cassette, see FIG. 2). Following oligonucleotide annealing, the modified plasmid was digested with SalI resulting in loss of the PAGL7PBP XhoI-SalI fragment and resulting in a vector/gene fragment.

The final step in the construction of the PDGF-B expression cassette was the ligation of the gene fragment into the vector/gene fragment to give plasmid pL7PBP (pAGL7PBP), as shown in FIG. 5. The PstI-BamHI insert fragment was isolated and nucleotide sequencing confirmed that the desired construction had been obtained. A map of the PDGF-B expression cassette in plL7PBP is shown in FIG. 4.

Construction of pYL7PBP (pYAGL7PBP)

The PDGF-B expression cassette of plL7PBP was isolated following BamHI digestion and inserted into the BamHI site of the yeast shuttle vector pAB24, described above, resulting in yeast expression plasmid pYL7PBP. A map of pYL7PBP is shown in FIG. 6. The nucleotide sequence of the complete expression cassette and the predicted amino acid sequence of the open reading frame (ORF) encoding truncated α-factor leader-Lys-Arg-proPDGF-B are given in SEQ ID NO: 35 and SEQ ID NO: 37, respectively. The complete nucleotide sequence of yeast expression plasmid pYL7PBP has been determined.

Expression Strain Identification: MB2-1(pYL7PBP)

Expression plasmid pYL7PBP was transformed into S. cerevisiae MB2-1 by standard procedures and plasmid-harboring, uracil prototrophs were selected as isolated colonies. Individual colonies from independent transformants were screened for expression following inoculation of isolated colonies into growth medium that selects for leucine prototrophs. The medium also is high in glucose to keep expression of sequences under ADR2 regulation (including the PDGF-B gene) repressed. Cultures were subsequently diluted and grown to confluence in low glucose, selective growth medium lacking uracil. Cell-free supernatants were assayed for PDGF-B by immunoactivity (ELISA) and by mitogenic activity on 3T3 cells. Frozen stocks were prepared of several transformants exhibiting consistently high levels of expression. Following repeated testing, the transformant exhibiting, on over age, the highest expression of PDGF-B, MB2-1 (pYL7PBP #22) was selected.

Example 4

Expression Plasmid pYJST400

The Lys-Arg dibasic processing site between the α-factor leader sequence and the N-terminal propeptide was elimi-
nated from expression plasmid pYL7PPB by in vitro mutagenesis to construct expression plasmid pYIST400. This pYIST400 has a single dibasic processing site, which resides at the preprotease/PGDF-B junction. Elimination of this first processing site was done to determine its relative effect on secretion of rhPGDF-BB from yeast as mediated by the α-factor leader peptide.

**Example 5**

Expression of Recombinant Human PDGF-BB

Recombinant human PDGF-BB is produced by a strain of the yeast, Saccharomyces cerevisiae, genetically modified with a multicopy yeast express ion plasmid that includes a gene encoding human PDGF-B. The preferred S. cerevisiae strain MB2-1 has the genotype: Matα ura3Δ, leu2–3, leu2–112, his3–11, his3–15, pep4Δ, [cir]. It is auxotrophic for uracil, leucine, and histidine, requiring these nutritional supplements when grown in minimal medium. MB2-1 does not contain an endogenous 2-μm plasmid, which tends to interfere with the stability of the introduced plasmids and encourages recombination between endogenous and introduced plasmids. The strain does not express functional protease A, the product of the PEPA gene, which interferes with the production of heterologous proteins. MB2-1 was designed to impart these favorable characteristics, which include selection for high expression of heterologous proteins.

Yeast expression plasmids pYAGL7PB, pYL7PPB, and pYIST400 were transformed into yeast strain MB2-1 as described by Hinnen et al. (1978) *Proc. Natl. Acad. Sci. USA* 75:1929–1933 and plated on ura-, glucone, sorbitol plates. Transformants were grown in leu-, 8% glucose liquid medium for 24 hours and then plated onto leu-, 8% glucose sorbitol plates to get individual colonies. Individual colonies were picked and grown in 3 ml of leu-, 8% glucose medium for 24 hours at 30°C, and then inoculated (1:50) into 1 liter of ura-, 1% glucose media and grown for 75 hours at 30°C. Yeast culture medium was assayed for PDGF activity by the human foreskin fibroblast mitogen assay (see Example 5 below).

As shown in Table 1, inclusion of the sequence encoding the N-terminal preprotease resulted in a mean 3.4-fold increase in secretion of rhPGDF-BB as measured by bioactivity and by ELISA. Additionally, elimination of the Lys-Arg processing site at the leader/preprotease junction resulted in a 2.8-fold decrease in rhPGDF-BB secretion (Table 1).

These results indicate that the presence of the native N-terminal preprotease enhances secretion of biologically active mature rhPGDF-BB when flanked by preferred processing sites that have been modified for improved recognition by a proteolytic enzyme of the yeast host cell. Thus, cleavage at the leader/preprotease junction, as well as at the propeptide/PGDF-B junction, apparently facilitates the proper folding and/or processing and/or transport of the pro-PGDF-B, resulting in enhanced secretion of mature rhPGDF-BB.

**Example 6**

Human Foreskin Fibroblast (HFF) Mitogen Assay for PDGF

Human foreskin fibroblast stocks were stored frozen; freezing was at passage 13. Prior to use, HFF were thawed and then grown in T75 flasks until confluent, which usually occurred at 5–7 days. Growth medium contained Dulbecco's Modified Eagles Medium (DMEM), 20% fetal bovine serum (FBS), 1 mM sodium pyruvate, 300 μg/ml L-glutamine, 100 U/ml penicillin, and 100 μg/ml streptomycin. Cells were incubated at 37°C in humidified 7% CO₂, 93% air atmosphere. At confluency, cells were passaged by rinsing the monolayer with phosphate buffered saline (PBS) lacking C₂⁺ and Mg²⁺, dissociating them in trypsin containing EDTA, and diluting them with growth medium. Cells were passaged no more than 8 times after thawing.

To assay for PDGF, HFFs were plated as follows. The cells were rinsed and dissociated with trypsin as above. The trypsinized cells were pelleted and resuspended to a concentration of 1×10⁵ cells/ml in medium similar to growth medium, except that 5% FBS replaced 20% FBS; 100 μl of suspension was dispensed into each well of a 96-well microtiter plate; and then the cells were incubated 5–6 days under the above described conditions.

PDGF in the sample was determined by monitoring [3H]thymidine incorporation into HFF DNA stimulated by PDGF. Samples were added to the wells containing HFF monolayers, and the assay plates incubated as above for 18 hours. The HFF cultures were then pulsed with [Methyl-³H]thymidine (10 μCi/ml final concentration, 1 μCi/well) at 37°C under the above described incubation conditions for 8 hours. After incubation, the cells were rinsed with PBS and fixed. Fixing was by incubation with 5% trichloroacetic acid (TCA) and then 100% methanol for 15 minutes, followed by drying in air. The cells were then solubilized with 0.3N NaOH and then counted in a liquid scintillation counter.

Control samples were treated as the samples described above and were prepared as follows. For positive controls, PDGF, purchased from PDGF, Inc., was dissolved to a final concentration of 100 ng/ml in DMEM containing 10 mg/ml BSA. A standard curve was prepared; the first point was 10 ng/ml, the remaining points were 2-fold serial dilutions. Each dilution was tested in triplicate. Negative controls, which lacked both sample and control PDGF, were also run.

**Example 7**

Expression Plasmids pYUI

Plasmid pYULUIGCF includes an expression cassette with the hybrid yeast promoter ADHI/GAP and Matα factor leader sequences fused to a sequence encoding the human IGF-I-A gene. This sequence was synthetically derived using yeast preferred codons. A dibasic amino acid processing site is present at the α-factor leader/IGF-I-A junction. The nucleotide sequence and predicted amino acid sequence of the open reading frame encoding α-factor leader/IGF-I-A primary translation product are given in SEQ ID NO: 38 and SEQ ID NO: 39, respectively.

Plasmid pYULUIGCF34 differs from pYULUIGCF4 only in its open reading frame. This cassette includes an open reading frame encoding a full length Matα factor leader sequence fused to a sequence encoding the human IGF-I-A gene with its C-terminal prosequence. Dibasic amino acid processing sites are present at the α-factor leader/IGF-I-A and IGF-I-A/IGF-I-A prosequence junctions. The nucleotide sequence and predicted amino acid sequence of the open reading frame encoding α-factor leader/proIGF-I-A primary translation product are given in SEQ ID NO: 40 and SEQ ID NO: 41, respectively.

Both of these plasmids were generated by inserting the respective expression cassette into the unique BamHI cloning site of the yeast shuttle vector pAB24 as described above.
Expression of Recombinant Human IGF-I-A

Recombinant human IGF-I-A is produced by a strain of the yeast *Saccharomyces cerevisiae*, genetically modified with a multicopy yeast expression plasmid that includes a gene encoding human IGF-I-A. Yeast expression plasmids pYLUIGF24 and pYLUIGF34 were transformed into a yeast strain by procedures previously mentioned.

Western blot data indicated that properly processed IGF-I-A protein was obtained with the prosequence, modified KEX2 processing site, and a yeast secretion leader.

All publications and patent applications mentioned in the specification are indicative of the level of those skilled in the art to which this invention pertains. All publications and patent applications are herein incorporated by reference to the same extent as if each individual publication or patent application was specifically and individually indicated to be incorporated by reference.

Although the foregoing invention has been described in some detail by way of illustration and example for purposes of clarity of understanding, it will be obvious that certain changes and modifications may be practiced within the scope of the appended claims.

### SEQUENCE LISTING

<table>
<thead>
<tr>
<th>ATG</th>
<th>AGA</th>
<th>TTT</th>
<th>CCT</th>
<th>TCA</th>
<th>ATT</th>
<th>TTT</th>
<th>ACT</th>
<th>GCA</th>
<th>GTT</th>
<th>TTG</th>
<th>TTC</th>
<th>GCC</th>
<th>TCG</th>
<th>ACC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Met</td>
<td>Arg</td>
<td>Phe</td>
<td>Pro</td>
<td>Ser</td>
<td>Ile</td>
<td>Phe</td>
<td>Thr</td>
<td>Ala</td>
<td>Val</td>
<td>Leu</td>
<td>Phe</td>
<td>Ala</td>
<td>Ser</td>
<td>Ser</td>
</tr>
<tr>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td>7</td>
<td>8</td>
<td>9</td>
<td>10</td>
<td>11</td>
<td>12</td>
<td>13</td>
<td>14</td>
<td>15</td>
</tr>
<tr>
<td>-17</td>
<td>-16</td>
<td>-15</td>
<td>-14</td>
<td>-13</td>
<td>-12</td>
<td>-11</td>
<td>-10</td>
<td>-9</td>
<td>-8</td>
<td>-7</td>
<td>-6</td>
<td>-5</td>
<td>-4</td>
<td>-3</td>
</tr>
</tbody>
</table>

Sequence description: SEQ ID No: 1;
GTT GAA GTT CAA AGA TGT TCT GGT TGG TGT AAC AAG AAC AAG GTT CAA
Val Glu Val Gin Arg Cys Ser Gly Cys Asn Asn Arg Asn Val Gin
45 50 55
TGT AGA CCA ACT CAA GTT CAA TGG AGA CCA GTT AGA AAG ATC
Cys Arg Pro Thr Gin Val Gin Leu Arg Pro Val Gin Val Arg Lys Ile
60 65 70 75
GAA ATC GTT AGA AAG CCA ATC TTC AAG AAG GTT ACT GTT ACT TGG
Glu Ile Val Arg Lys Pro Ile Phe Lys Lys Ala Thr Val Thr Leu
80 85 90
GAA GAC CAC TGG GTT TGG TAA GAA GTT AAG TGG AAG ACT GTC GCC GCC AGG CCA
Glu Asp His Leu Ala Cys Lys Cys Glu Thr Val Ala Ala Ala Arg Pro
95 100 105

GTT ACT TAA TAG
Val Thr *
110

(2) INFORMATION FOR SEQ ID NO: 2:

(i) SEQUENCE CHARACTERISTICS:
   (A) LENGTH: 146 amino acids
   (B) TYPE: amino acid
   (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 2:

Met Arg Phe Pro Ser Ile Phe Thr Ala Val Leu Phe Ala Ala Ser Ser
-37 -35 -33 -30 -25
Ala Leu Ala Ala Pro Val Asn Thr Thr Thr Glu Asp Glu Thr Ala Gin
-29 -27 -25 -20 -10
Ile Pro Ala Lys Arg Ser Leu Gly Ser Leu Thr Ile Ala Glu Pro Ala
-5 1 5 10
Met Ile Ala Glu Cys Lys Thr Arg Thr Glu Val Phe Glu Ile Ser Arg
15 20 25
Arg Leu Ile Asp Arg Thr Asn Ala Asn Phe Leu Val Thr Pro Pro Cys
30 35 40
Val Glu Val Gin Arg Cys Ser Gly Cys Asn Asn Arg Asn Val Gin
45 50 55
Cys Arg Pro Thr Gin Val Gin Leu Arg Pro Val Gin Val Arg Lys Ile
60 65 70 75
Glu Ile Val Arg Lys Pro Ile Phe Lys Lys Ala Thr Val Thr Leu
80 85 90
Glu Asp His Leu Ala Cys Lys Cys Glu Thr Val Ala Ala Ala Arg Pro
95 100 105
Val Thr

(2) INFORMATION FOR SEQ ID NO: 3:

(i) SEQUENCE CHARACTERISTICS:
   (A) LENGTH: 352 base pairs
   (B) TYPE: nucleic acid
   (C) STRANDEDNESS: single
   (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(vi) ORIGINAL SOURCE:
   (A) ORGANISM: Homo sapiens

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 3:

CTCATGATAA AAGAATCTTTG GOTTCTTTGA CTATCGCTGA ACCAGCTATG ATGCTGUAT
GTAGACCTAG AACTGAAGTT TTCGAAATCT CCAGAAGATT ACTTCTTGGT GAAACGTTCA TTGGCCACCA TGTGTTGAAG ATGTAGACCA, ACTCAAGTTC TTCAAAGATG AATTGAGACC AAATCGTTAG AAAGAAGCCA ATCTTCAAGA AGGCTACTGT CTTGTAAGTG

(2) INFORMATION FOR SEQ ID NO: 4:

(i) SEQUENCE CHARACTERISTICS:
   (A) LENGTH: 352 base pairs
   (B) TYPE: nucleic acid
   (C) STRANDEDNESS: single
   (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid
   (vi) ORIGINAL SOURCE: Homo sapiens

(xi). SEQUENCE DESCRIPTION: SEQ ID NO: 4: TCGACGACGC GTGGTCTTCC CTTTCTAACT GTTACAACAA GTTAGTTCTG AGCGATCATA

5

GTGCTCTCC AAAAGACAG TAGCCTTTCT GGAGATTGCG TTCCTTTCAA CGATTTGCTAT

CTTCCTAATC TCAATGCTG TCAATGCTG CTATTTTGCT CTATTTTGCT CTATTTTGCT

GTAAAACAA CCAGAACACT TTTGAACTC AAGCGATGCT GCCTAACCAC AGAAGTACCC

GTGAGCTGCTTGAACTAG CTTCCGAGA TTGCGAAAC TTGAGCTCTAG CTCTACATCC

AGGATGATAC GCTGCTTCTG CTTCTCTGAA AGGATGTTGA GTTCTTTTAT CT

(2) INFORMATION FOR SEQ ID NO: 5:

(i) SEQUENCE CHARACTERISTICS:
   (A) LENGTH: 44 base pairs
   (B) TYPE: nucleic acid
   (C) STRANDEDNESS: single
   (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA
   (vi) ORIGINAL SOURCE: Homo sapiens

(xi). SEQUENCE DESCRIPTION: SEQ ID NO: 5: GCTATGATCG CTGAATGTAA GACTAGAACT GAAGTCTTGC AAATC

(2) INFORMATION FOR SEQ ID NO: 6:

(i) SEQUENCE CHARACTERISTICS:
   (A) LENGTH: 45 base pairs
   (B) TYPE: nucleic acid
   (C) STRANDEDNESS: single
   (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA
   (vi) ORIGINAL SOURCE: Homo sapiens

(xi). SEQUENCE DESCRIPTION: SEQ ID NO: 6: GCTATCAGCT GTGAAGTCTTGC AAATC

(2) INFORMATION FOR SEQ ID NO: 7:

(i) SEQUENCE CHARACTERISTICS:
   (A) LENGTH: 45 base pairs
   (B) TYPE: nucleic acid
   (C) STRANDEDNESS: single
   (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA
   (vi) ORIGINAL SOURCE: Homo sapiens

(xi). SEQUENCE DESCRIPTION: SEQ ID NO: 7: GCTATGAGC GTGAAGTCTTGC AAATC
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 45 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 7:
TCCAGAAGATGATCGACAGAACTAACGCT
AACTTCCTGGTTTTGG

(2) INFORMATION FOR SEQ ID NO: 8:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 45 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 8:
CCACCATGTGTTGAAGTCAAAAGATGTTCTGGTTGTTGTAACAAC

(2) INFORMATION FOR SEQ ID NO: 9:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 45 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 9:
AGAAACGTTCAAATGTAGACCAACTCAAGTTCAATTGAGAC
CAGTT

(2) INFORMATION FOR SEQ ID NO: 10:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 45 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 10:
CAAGTTAGAAAGATCGAAATCGTTAGAAAGAAGCCAATCTTCAAG

(2) INFORMATION FOR SEQ ID NO: 11:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 44 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA
(vi) ORIGINAL SOURCE:  
(A) ORGANISM: Saccharomyces cerevisiae

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 11:

AAAGGCTACTG TTACTTTGGA AGACCACTTG GCTTGTAAGT, GTGA

(2) INFORMATION FOR SEQ ID NO: 12:

(i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 39 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(vi) ORIGINAL SOURCE:  
(A) ORGANISM: Homo sapiens

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 12:

AACTGTTGCT GGTGCTAGAC CAGTTACTTA ATAGCGTCG

(2) INFORMATION FOR SEQ ID NO: 13:

(i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 18 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(vi) ORIGINAL SOURCE:  
(A) ORGANISM: Homo sapiens

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 13:

TCTATTTTCT AGAAACCC

(2) INFORMATION FOR SEQ ID NO: 14:

(i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 45 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(vi) ORIGINAL SOURCE:  
(A) ORGANISM: Homo sapiens

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 14:

AAGAAACTGA TAGCGACTTG GTCGATACTA. GCGACTTACA. TTCTG

(2) INFORMATION FOR SEQ ID NO: 15:

(i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 45 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(vi) ORIGINAL SOURCE:  
(A) ORGANISM: Homo sapiens

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 15:

ATCTTGACTT CAAAAGCTTT AGAGGTCTTC TAACTAGCTG TCTTG
(2) INFORMATION FOR SEQ ID NO: 16:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 45 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 16:

ATCCGATGG AGAACACCA CGCGTTTAC ACAACTTCAA GTTTC

(2) INFORMATION FOR SEQ ID NO: 17:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 45 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 17:

TACAAGACCA ACAACATTGT TGTCTTTGCA AGTTACATCT GGTTG

(2) INFORMATION FOR SEQ ID NO: 18:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 45 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 18:

AGTTCAAGTT AACTCTGGTC AAGTTCAATC TTTCTAGCTT TAGCA

(2) INFORMATION FOR SEQ ID NO: 19:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 44 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 19:

ATCTTTCTTC GGTTAGAAGT TCTTCCGATG ACAATGAAAC CTTC

(2) INFORMATION FOR SEQ ID NO: 20:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 44 base pairs
(B) TYPE: nucleic acid
(ii) MOLECULE TYPE: cDNA

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 20:

TGCTGACG AACATCCA CTTGACAC GACGACGATC TGGT

---

(ii) MOLECULE TYPE: cDNA

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 21:

CAATGAATTA TGGCAGCAGC

---

(ii) MOLECULE TYPE: other nucleic acid
(A) DESCRIPTION: /desc = "Assembled synthetic"

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Synthetic (derived from Saccharomyces cerevisiae)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 22:

TTTTATCC AGCCCGACG GAATAGCTG CTCAGCTCA CACTACAACA GAAGATGAAA CGGCACAAAT TCCGGCTAAA

---

(ii) MOLECULE TYPE: other nucleic acid
(A) DESCRIPTION: /desc = "This sequence is the"

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Synthetic (derived from Saccharomyces cerevisiae)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 23:

GATCTTTTAG CCGGAATTTG TGCCGTTTCA TCTTCTGTTG TAGTGTTGAC TGGAGCAGCT AATGCGCTCG AGGCTGCGAA TAAAACTGCA

---

(ii) MOLECULE TYPE: cDNA
(A) LENGTH: 1845 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid
(A) DESCRIPTION: /desc = "Synthetic chimera"

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens/Saccharomyces cerevisiae

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 1115..1558

(ix) FEATURE:
(A) NAME/KEY: promoter
(B) LOCATION: 1115..1225
(D) OTHER INFORMATION: /standard_name= "ADH/GAP promoter"

(ix) FEATURE:
(A) NAME/KEY: misc_feature
(B) LOCATION: 1115..1225
(D) OTHER INFORMATION: /function= "mediates secretion of rhPDGF-B"
/product= "truncated alpha factor leader/signal peptide"
/standard_name= "alpha factor leader/signal sequence"

(ix) FEATURE:
(A) NAME/KEY: mat_peptide
(B) LOCATION: 1226..1558
(D) OTHER INFORMATION: /product= "rhPDGF-B peptide"
/standard_name= "rhPDGF-B"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 24:

GOATCCTTCA ATATGCGCAC ATACCTGTGT ATGTTCAAGG TCCCTCTGTT TAAGAAGCAA 60
AGCGCCCTTC GGGTTTGAAG ATGTTCAAGG TGGTCAAAT CTAAGAAATT TGGAATCC
CAGTCGTTAG CTAGCTAAGT ATACCAATGG CAAAGCGAG ACAACAAATAC CAGTCGCGAT 120
CAACGCGCA CACTCCTCCC GTAAGCTTCT CTAAGTTTCT TCTCGAGATG ATGGGAGAT
TACACACCT AAGCCTGGG AAAGTGTATT TCGGGATTCT TACACCAACTC CTACACAGAA 180
TGGTCGACAT GAGGCGTACT TGGACCAAGA TCTCGACCTC TTCTGCAATA TTGGGAGAAA 240
TACACCCGCT AGTTCAATATT AGGGGACATT GGGGAAAAC AGAACCGACC ACAGAAGGAA 300
ATCAATACTT AACAGTGAGA TGAGGGGACATT GGAGGAAAAC AGAACCGACC ACAGAAGGAA 360
TTAGAGTTTC TCAATTTCT CAACAGCAAA AGAACCGACC ACAGAAGGAA 420
ATCAGGTCAC TTAATGACAA GGGGAAAAC AGAACCGACC ACAGAAGGAA 480

GOATCCTTCA ATATGCGCAC ATACCTGTGT ATGTTCAAGG TCCCTCTGTT TAAGAAGCAA 60
AGCGCCCTTC GGGTTTGAAG ATGTTCAAGG TGGTCAAAT CTAAGAAATT TGGAATCC

TTAGAGTTTC TCAATTTCT CAACAGCAAA AGAACCGACC ACAGAAGGAA 660
TTAGAGTTTC TCAATTTCT CAACAGCAAA AGAACCGACC ACAGAAGGAA 720
TTAGAGTTTC TCAATTTCT CAACAGCAAA AGAACCGACC ACAGAAGGAA 780
TTAGAGTTTC TCAATTTCT CAACAGCAAA AGAACCGACC ACAGAAGGAA 840
TTAGAGTTTC TCAATTTCT CAACAGCAAA AGAACCGACC ACAGAAGGAA 900
TTAGAGTTTC TCAATTTCT CAACAGCAAA AGAACCGACC ACAGAAGGAA 960
TTAGAGTTTC TCAATTTCT CAACAGCAAA AGAACCGACC ACAGAAGGAA 1020
TTAGAGTTTC TCAATTTCT CAACAGCAAA AGAACCGACC ACAGAAGGAA 1080

GOATCCTTCA ATATGCGCAC ATACCTGTGT ATGTTCAAGG TCCCTCTGTT TAAGAAGCAA 1132
AGCGCCCTTC GGGTTTGAAG ATGTTCAAGG TGGTCAAAT CTAAGAAATT TGGAATCC

TTG ACT GCA GTT TTA TTC GCA GCC TCG AGC GCA TTA GCT GCT GCA GTC
GGAATACACAT CTAATGACAA GGGGAAAAC AGAACCGACC ACAGAAGGAA 1180

Phe Thr Ala Val Leu Phe Ala Ser Ser Ser Ala Leu Ala Ala Pro Val
<table>
<thead>
<tr>
<th>AUG ACT ACA GAA GAT GAA AGC CAA ATT CCG GCC ACAA GAA TGT</th>
<th>1228</th>
</tr>
</thead>
<tbody>
<tr>
<td>Asn Thr Thr Thr Glu Asp Glu Thr Ala Gln Ile Pro Ala Lys Arg Ser</td>
<td>-15</td>
</tr>
<tr>
<td>TTG GTT CTC ACT ATC GCT GAA CCA GCT ATG ATC GCT GAA TGT AAG</td>
<td>1276</td>
</tr>
<tr>
<td>Leu Gly Ser Leu Thr Ile Ala Glu Pro Ala Met Ile Ala Glu Cys Lys</td>
<td>5</td>
</tr>
<tr>
<td>ACT AGA ACT GAA GTT TTC CCA AGA CCC GAA GTT CAA AGA TTC</td>
<td>1324</td>
</tr>
<tr>
<td>Thr Arg Thr Glu Val Phe Glu Ile Ser Arg Arg Leu Ile Asp Arg Thr</td>
<td>20</td>
</tr>
<tr>
<td>AAC GCT AAC TTC TTG GTG CCA CCA TGT GAA GTT CAA AGA TGT</td>
<td>1372</td>
</tr>
<tr>
<td>Asn Ala Asn Phe Leu Val Trp Pro Cys Val Val Glu Val Gin Cys</td>
<td>35</td>
</tr>
<tr>
<td>TCT GCT CAC AGA AAG CAC GTA AGT ACA ACT GAA TTC</td>
<td>1420</td>
</tr>
<tr>
<td>Ser Gly Cys Asn Asn Arg Asp Val Glu Cys Arg Pro Thr Gin Val</td>
<td>50</td>
</tr>
<tr>
<td>CAA TTG AGA GTA CAA GAA AGA AGT CAA ACT CAA GTT</td>
<td>1468</td>
</tr>
<tr>
<td>Gin Leu Arg Pro Val Gin Val Gin Arg Arg Thr CAA ACT CAA GTT</td>
<td>70</td>
</tr>
<tr>
<td>GCC GTA CAA GAA AGA AGA TTC CCA GAA TTC</td>
<td>1516</td>
</tr>
<tr>
<td>Pro Ile Phe Leu Pro Val Thr Val Thr Leu Leu Arg His Leu Ala Cys</td>
<td>95</td>
</tr>
<tr>
<td>AAG TGT GAA ACT GCC GCC AGC CAA GTT ACT TAA TAG</td>
<td>1558</td>
</tr>
<tr>
<td>Lys Cys Glu Thr Val Ala Ala Arg Pro Val Thr   *   *</td>
<td>100</td>
</tr>
<tr>
<td>CCG GCT AAA AGA TCT Pro Ala Lys Arg Ser</td>
<td>1276</td>
</tr>
<tr>
<td>ATC GCT TCT CCA ACC AGA AGA ACC AGT ACT CAA ACT CAA GTT</td>
<td>1420</td>
</tr>
<tr>
<td>Ser Gly Cys Asn Asn Arg Asp Val Glu Cys Arg Pro Thr Gin Val</td>
<td>50</td>
</tr>
<tr>
<td>CAA TTG AGA GTA CAA GAA AGA AGT CAA ACT CAA GTT</td>
<td>1468</td>
</tr>
<tr>
<td>Gin Leu Arg Pro Val Gin Val Gin Arg Arg Thr CAA ACT CAA GTT</td>
<td>70</td>
</tr>
<tr>
<td>GCC GTA CAA GAA AGA AGA TTC CCA GAA TTC</td>
<td>1516</td>
</tr>
<tr>
<td>Pro Ile Phe Leu Pro Val Thr Val Thr Leu Leu Arg His Leu Ala Cys</td>
<td>95</td>
</tr>
<tr>
<td>AAG TGT GAA ACT GCC GCC AGC CAA GTT ACT TAA TAG</td>
<td>1558</td>
</tr>
<tr>
<td>Lys Cys Glu Thr Val Ala Ala Arg Pro Val Thr   *   *</td>
<td>100</td>
</tr>
</tbody>
</table>

**Information for Seq ID No: 25**

- **Sequence Characteristics:**
  - **Length:** 146 amino acids
  - **Type:** amino acid linear
  - **Topology:** Linear

- **Molecule Type:** Protein

**Sequence Description:**

<table>
<thead>
<tr>
<th>Met Arg Phe Pro Ser Ile Phe Thr Ala Val Leu Phe Ala Ala Ser Ser</th>
<th>-37</th>
</tr>
</thead>
<tbody>
<tr>
<td>Asn Thr Thr Thr Glu Asp Glu Thr Ala Gln Ile Pro Ala Lys Arg Ser</td>
<td>-15</td>
</tr>
<tr>
<td>TTG GTT CTC ACT ATC GCT GAA CCA GCT ATG ATC GCT GAA TGT AAG</td>
<td>1276</td>
</tr>
<tr>
<td>Leu Gly Ser Leu Thr Ile Ala Glu Pro Ala Met Ile Ala Glu Cys Lys</td>
<td>5</td>
</tr>
<tr>
<td>ACT AGA ACT GAA GTT TTC CCA AGA CCC GAA GTT CAA AGA TTC</td>
<td>1324</td>
</tr>
<tr>
<td>Thr Arg Thr Glu Val Phe Glu Ile Ser Arg Arg Leu Ile Asp Arg Thr</td>
<td>20</td>
</tr>
<tr>
<td>AAG TGT GAA ACT GCC GCC AGC CAA GTT ACT TAA TAG</td>
<td>1558</td>
</tr>
<tr>
<td>Lys Cys Glu Thr Val Ala Ala Arg Pro Val Thr   *   *</td>
<td>100</td>
</tr>
</tbody>
</table>

**Sequence Description:**

- **Seq ID No:** 25
- **Length:** 146 amino acids
- **Type:** Amino Acid Linear
- **Topology:** Linear

**Molecule Type:** Protein
Glu Asp His Leu Ala Cys Lys Cys Glu Thr Val Ala Ala Ala Arg Pro
92 100 105
Val Thr

(2) INFORMATION FOR SEQ ID NO: 26:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 621 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid
(A) DESCRIPTION: /desc = "This construct is a

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Saccharomyces cerevisiae/Homo sapiens

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 1..621

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 26:

ATG AGA TTT CCT TCA ATT TTT ACT GCA GGT TTA TTC GCA GGC TCG AGC
Met Arg Phe Pro Ser Ile Phe Thr Ala Val Leu Phe Ala Ala Ser Ser
-96 -95 -90 -85

GCA TTA GCT GCT CCA GTG AAC ACT ACA ACA GAA GAT GAA ACG GCA CAA
Ala Leu Ala Ala Pro Val Aen Thr Thr Glu Aep Glu Thr Ala Gln
-80 -75 -70 -65

ATT CCC GCT ATT CCC GAG GAG CTC TAC GAG ATG CTG
Ile Pro Ala Lys Arg Aep Pro Ile Pro Glu Glu Leu Tyr Glu Met Leu
144

AGT GAC CAC TCG ATC CCC TCC TTT GAT GAT CTC CAA CCC CTG CAC
Ser Asp His Ser Ile Arg Ser Phe Aep Aep Leu Gln Arg Leu His
-40 -40 -35

GGA GCC GCA GAG GAA GAT GGG GCC GAG TGT GAC CTC AAC ATG ACC
Gly Aep Pro Gly Glu Leu Aep Glu Ala Leu Aep Leu Met Thr
-30 -25 -20 -20

GCC TCC CAC TCT GCA GCC GAG CTC GAG ACC TGT CCT CGG GGG AAG AGG
Arg Ser His Ser Gly Gly Leu Glu Ser Leu Ser Leu Ala Arg Lys Arg
-15 -10 -5

ACC CTC GGT GTC TCC CTG ACC ATT GCT GAG CCG GGC ATC GCC GAG TCC
Ser Leu Gly Ser Leu Thr Ile Ala Glu Pro Ala Met Ile Ala Ala Cys
336

AAG ACC CCC ACC GAG GTG TCC GAG ATC TCC CGG CTC ATC GCC GAC CCC
Lys Thr Arg Thr Glu Val Phe Glu Ile Ser Arg Arg Leu Ile Aep Arg
384
| ACC AAC GCC AAC TTC CTO GTO TGG CCC TOT GTO GAG GTO CAG CUC | Thr Asn Ala Asn Phe Leu Val Trp Pro Pro Cys Val Glu Val Gin Arg |
|---------------------------------------------------------------|
| 35 40 45                                                     |
| TGC GCG TGC TGC AAC AAC CCG TGC CAG CGG CCC ACC CAG          |
| 50 55 60                                                     |
| Cys Ser Gly Cys Asn Asn Arg Asn Val Gin Cys Arg Pro Thr Gin   |
| 65 70 75 80                                                  |
| GTG CAG CTG CGA CCT GTC CAG GTO AGA AGA ATG GTC CGG CAG       |
| 85 90                                                       |
| Val Gin Leu Arg Pro Val Gin Val Arg Lys Ile Glu Ile Val Arg Lys |
| 95 100                                                      |
| AAG CCA ATC TTT AAG AAG GCC AGC GTC AGC CTG GAA GAC CAC CTG CCA |
| 100 105                                                     |
| Lys Pro Ile Phe Lys Ala Thr Val Thr Leu Gin Asp His Leu Ala   |
| 110 115                                                     |
| TGC AAG TGT GAG ACA GTC GCA GCT GCA CGT GTC ACC TAA TAG       |
| 120 125                                                     |
| Cys Lys Cys Gin Thr Val Ala Ala Ala Arg Pro Val Thr           |

(2) INFORMATION FOR SEQ ID NO: 27:

(i) SEQUENCE CHARACTERISTICS:
   (A) LENGTH: 205 amino acids
   (B) TYPE: amino acid
   (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(x) SEQUENCE DESCRIPTION: SEQ ID NO: 27:

Met Arg Phe Pro Ser Ile Phe Thr Ala Val Leu Phe Ala Ala Ser Ser
-96 -95 -90 -85
Ala Leu Ala Ala Pro Val Asn Thr Thr Gin Gin Gin Gin Gin Gin Gin Gin
-90  -75  -70  -65
Ile Pro Ala Lys Arg Asp Pro Ile Pro Gin Gin Gin Gin Gin Gin Gin Gin
-60  -55  -50
Ser Asp His Ser Ile Arg Ser Phe Asp Gin Gin Gin Gin Gin Gin Gin Gin
-45  -40  -35
Gly Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin
-30  -25  -20
Arg Ser His Ser Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin
-15  -10  -5
Ser Leu Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin
  5  10  15
Lys Thr Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin
  20  25  30
Thr Asn Ala Asn Phe Leu Val Thr Pro Pro Cys Gin Gin Gin Gin Gin
  35  40  45
Cys Ser Gin Cys Asn Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin
  50  55  60
Val Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin
  65  70  75  80
Lys Pro Ile Phe Lys Ala Thr Val Thr Leu Gin Asp His Leu Ala
  85  90  95
Cys Lys Cys Gin Thr Val Ala Ala Ala Arg Pro Val Thr
  100 105

(2) INFORMATION FOR SEQ ID NO: 28:

(i) SEQUENCE CHARACTERISTICS:
   (A) LENGTH: 1320 base pairs
   (B) TYPE: nucleic acid
   (C) STRANDINESS: single
(D) TOPOLGY: linear

(ii) MOLECULE TYPE: other nucleic acid
(A) DESCRIPTION: /desc = "Synthetic chimera"

(vi) ORIGIN: Homo sapiens/Saccharomyces cerevisiae

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 454..1179

(ix) FEATURE:
(A) NAME/KEY: misc_feature
(B) LOCATION: 454..519
(D) OTHER INFORMATION: /product= "PDGF-B prepeptide"
/standard_name= "PDGF-B presequence"

(ix) FEATURE:
(A) NAME/KEY: transit_peptide
(B) LOCATION: 455..696
(D) OTHER INFORMATION: /function= "mediates protein transport"
/product= "PDGF-B prepeptide"
/standard_name= "PDGF-B presequence"

(ix) FEATURE:
(A) NAME/KEY: met_peptide
(B) LOCATION: 697..1023
(D) OTHER INFORMATION: /product= "rhPDGF-B peptide"
/standard_name= "rhPDGF-B"

(ix) FEATURE:
(A) NAME/KEY: transit_peptide
(B) LOCATION: 1024..1179
(D) OTHER INFORMATION: /function= "mediates protein transport"
/product= "PDGF-B prepeptide"
/standard_name= "PDGF-B presequence"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 28:

GAATTCCAC AAAAAATCC AAAAAACTA AGCCGCGCGG CAGAGGAAAA CGCCTUTAGC 60
CGCCGACTGA AGACAGCAAC ACACTACCG TGTTCTCTCT CTCCTCAGAG GTGGAUTGTC 120
CCTGGCGGCG CCGACACGCG TAGACGCTCG GCTGTTGCTTG CCAACGACCC CCGGCGCGTCOT 180
GGATCTGCA CTGCGGCCGG GATCGCGGCG CGTGCGCCCA GCTGCTCGCCG CGTGCTGCTCC 240
CAGCGCTGCT CGCTCGCCCG CAGCGCCAGA GCGCGCGGCG GCGCGCGGCG GCGCGCGCG 300
ATGCGGCGTA GCGCGCGCGC CAGCGCCGCTG AGCGCGCGAT CCAGCGCGAG CCGCGCGCG 360
CCACCGCCGC CCACCGCGCG CCACCGCGCG CGCGCGGTGC ATCTACGCGTCAGCTC 420
TGCGCGCGCG GCGCGGCGCG CCGGGGCTGC GCCATGTGCC TGGTGGCG CGT 474

TTG CTG CTG CTG CTG AC CGG GAG GAG GAC Phe Leu Ser Leu Cys Tyr Leu Arg Leu Val Ser Ala Gly Aep -70 -65 -60

CCC ATT CCC GAG GAG CTG TTG ACC AAC GCC AAC GAG GAG AAC Pro Ile Pro Glu Glu Leu Tyr Glu Met Leu Ser Aep His Ser Ile Arg -55 -50 -45

TCC TTT GAT GAT CTC CAA GCC CTG CGC CTG AAC GHA GAC CCC GGA GAG GAA Ser Phe Aep Leu Gin Arg Leu His Gly Aep Pro Gly Glu Glu -40 -35 -30

GAT GGG GCC GAG TTG GAG CTG AAC ATG ACC CCC TAC TCT GTA GGC Aep Gly Ala Glu Leu Aep Leu Asn Met Thr Arg Ser His Ser Gly Gly Thr -23 -20 -15

GAG CTG GAG ACC TTG GCT CTT GTA GAG AGA AGG ACC CTG GGT GCC TCC CTG ACC Glu Leu Glu Ser Leu Ala Arg Gly Arg Arg Ser Leu Thr -10 -5 -1 5
ATT OCT GAG CGG GCC ATG ATC GAG GCC ATG ACG CGC ACC ACC GAG GTG
Ile Ala Glu Pro Ala Met Ile Ala Glu Cys Lys Thr Arg Thr Glu Val
10 15 20

TTC GAG ATC TCC CGC GCC ATG AAG ACC GCC ACC GAG GTG
Phe Glu Ile Ser Arg Arg Leu Ile Asp Arg Thr Ann Ala Ann Phe Leu
25 30 35

GCU GCG CCG GGT GTG GAG GTC CAG GCC GCC GCC GCC ACC
Val Trp Pro Pro Cys Val Glu Val Glu Arg Ser Gly Cys Cys Ann
40 45 50

AAC GCC ACC GTC CAG TCC GCC ACC CAG GGT GTC GTC CTG CTG
Ann Arg Ann Val Glu Cys Arg Pro Thr Gin Val Glu Leu Arg Pro Val
55 60 65 70

GAT GTA AGA AAG ATC GAG ATT GTA CGG AAG CAG GAC ATA TTT AAG AGG
Gln Val Arg Ile Glu Ile Val Arg Lys Pro Ile Pro Lys Lys
75 80 85

GCC AGC GTG AGC CTG GAA GAC GAC CTG GAA TGG GTG GAG ACA GTC
Ala Thr Val Thr Leu Glu Asp His Leu Ala Cys Lys Glu Thr Val
90 95 100

GCA GTT GCA CGG CCT GTG ACC CGA AGC CGG GGT TCC CAG GAG CAG
Ala Ala Ala Arg Pro Thr Pro Thr Ser Pro Met Pro Thr Pro Thr
105 110 115

CGA GCC AAA ACC CCC CAA ACT GGT ACC ATT GCG GTG CGA GTG
Arg Ala Lys Thr Pro Gin Thr Arg Val Thr Arg Thr Val Arg Val
120 125 130

CGG CCG CCC AGG GCC AGG CAC CGG AAA TTC AAC ACC CAT GAC
Arg Arg Pro Pro Lys Lys His Arg Lys Phe Lys His Thr His Asp
135 140 145 150

AGA ACC GCA CTG AAG GAC ACC CTT GCA TCC TAG GGCATCGCC AGUGAAGT
Lys Thr Ala Leu Leu Tyr His Thr Gln Arg Cys Glu Arg Lys
155 160

GTGGCAAGG TTATTTAATA TGTATTGTGT ATTATCCGCT CCTGGGCTCCT GATGATGATGAT
1259

AAATATATTTCCCTCGTCGCTCG CTCTGCGTGC CGTGGCGTCA CGTGGCGTCC
1319

C

1320

(2) INFORMATION FOR SEQ ID NO: 29:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 241 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 29:

Met Asn Arg Cys Trp Ala Leu Phe Leu Ser Leu Cys Tyr Leu Arg
-81 0 -70
Leu Val Ser Ala Glu Gly Asp Pro Pro Gly Glu Leu Tyr Glu Met
-65 0 -55 -50
Leu Ser Asp His Ser Ile Arg Ser Phe Asp Asp Leu Gin Arg Leu Leu
-45 0 -40 -35
His Gly Asp Pro Gly Glu Asp Gly Ala Glu Leu Asp Leu Asp Met
-30 0 -25 -20
Thr Arg Ser His Ser Gly Glu Leu Glu Ser Leu Ala Arg Gly Arg
-15 0 -10 -5
Arg Ser Leu Gly Ser Leu Thr Ile Ala Glu Pro Ala Met Ile Ala Glu
1 0 5 10 15
Cys Lys Thr Arg Thr Glu Val Phe Glu Ile Ser Arg Arg Leu Ile Asp
20 25 30
Arg Thr Asn Ala Asn Phe Leu Val Trp Pro Pro Cys Val Glu Val Gln
45
Arg Cys Ser Gly Cys Cys Asn Asn Arg Asn Val Gln Cys Arg Pro Thr
50
Gln Val Gln Leu Arg Pro Val Gln Val Arg Lys Ile Glu Ile Val Arg
55
Lys Lys Pro Ile Phe Lys Lys Ala Thr Val Thr Leu Glu Asp His Leu
60
85
90
95
Ala Cys Lys Cys Glu Thr Val Ala Ala Ala Arg Pro Val Thr Arg Ser
100
105
110
Pro Gly Gly Ser Gln Glu Gln Arg Ala Lys Thr Pro Gln Thr Arg Val
115
120
125
Thr Ile Arg Thr Val Arg Val Arg Pro Pro Lys Gln Thr Arg Arg
130
135
140
Lys Phe Lys His Thr His Asp Lys Thr Ala Leu Lys Glu Thr Leu Gly
145
150
155
Ala
160

(2) INFORMATION FOR SEQ ID NO: 30:

(i) SEQUENCE CHARACTERISTICS:
   (A) LENGTH: 34 base pairs
   (B) TYPE: nucleic acid
   (C) STRANDEDNESS: double
   (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid
   (A) DESCRIPTION: /desc = "Synthetic oligonucleotide"

(vi) ORIGINAL SOURCE:
   (A) ORGANISM: Synthetic (derived from Homo sapiens sequence)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 30:

CATCCCGAG GAGCTCTACG AGATGCTGAG TGAC

(2) INFORMATION FOR SEQ ID NO: 31:

(i) SEQUENCE CHARACTERISTICS:
   (A) LENGTH: 27 base pairs
   (B) TYPE: nucleic acid
   (C) STRANDEDNESS: double
   (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid
   (A) DESCRIPTION: /desc = "Synthetic oligonucleotide"

(vi) ORIGINAL SOURCE:
   (A) ORGANISM: Synthetic (derived from Homo sapiens sequence)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 31:

CTTGGCTCGG GGGAAGAGGA GCCTGG

(2) INFORMATION FOR SEQ ID NO: 32:

(i) SEQUENCE CHARACTERISTICS:
   (A) LENGTH: 89 base pairs
   (B) TYPE: nucleic acid
   (C) STRANDEDNESS: single
   (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid
   (A) DESCRIPTION: /desc = "Synthetic oligonucleotide"
(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens/Saccharomyces cerevisiae derived sequence

(ix) FEATURE:
(A) NAME/KEY: misc_feature
(B) LOCATION: 44..89
(D) OTHER INFORMATION: /function= "truncated alpha factor leader/lys-arg proc./N-term. propept"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 32:
TCGACGCGAT TAGCTGCTCC AGTCAACACT ACAACAGAAG ATGAAACGGC ACAAATTCCG 60
GCTAAAAGAG ACCCCATTCC CGAGGAGCT 89

(2) INFORMATION FOR SEQ ID NO: 33:
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 81 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid
(A) DESCRIPTION: /desc = "Synthetic oligonucleotide"

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens derived sequence

(ix) FEATURE:
(A) NAME/KEY: misc_feature
(B) LOCATION: 1..39
(D) OTHER INFORMATION: /function= "C-term. alpha factor leader/lys-arg proc./N-term. propeptide"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 33:
CCTCGGGAAT GGGTCTCTT, TTAGCCGGAA TTTGTGCCGT TTCATCTTCT GTTGTAGTGT 60
TGACTGGAC AGCTAATGC C 81

(2) INFORMATION FOR SEQ ID NO: 34:
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 48 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid
(A) DESCRIPTION: /desc = "Synthetic oligonucleotide"

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens derived sequence

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 34:
CAAGTGTGAG ACAGTGCGAG CTGCGGCGGC TTGACCTAA TAGCTGC 48

(2) INFORMATION FOR SEQ ID NO: 35:
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 56 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid
(A) DESCRIPTION: /desc = "Synthetic oligonucleotide"

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens derived sequence

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 35:
US 6,706,496 B2

TCGACGACGC TATTAGGTCA CAGGCCGTGC AGCTGCCACT GTCTCACACT TGCATG

(2) INFORMATION FOR SEQ ID NO: 36:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 2023 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid
(A) DESCRIPTION: /desc = "Synthetic chimera"

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens/Saccharomyces cerevisiae

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 1119..1735

(ix) FEATURE:
(A) NAME/KEY: promoter
(B) LOCATION: 1..1114
(D) OTHER INFORMATION: /standard_name= "ADH/GAP promoter"

(ix) FEATURE:
(A) NAME/KEY: misc_feature
(B) LOCATION: 1115..1225
(D) OTHER INFORMATION: /function= "mediates secretion of rhPDGF-B"
/product= "alpha factor signal/truncated alpha factor leader peptide"
/standard_name= "truncated alpha factor signal/leader sequence"

(ix) FEATURE:
(A) NAME/KEY: transit_peptide
(B) LOCATION: 1226..1402
(D) OTHER INFORMATION: /product= "PDGF-B propeptide"
/standard_name= "PDGF-B prosequence"

(ix) FEATURE:
(A) NAME/KEY: mat_peptide
(B) LOCATION: 1403..1735
(D) OTHER INFORMATION: /product= "rhPDGF-B protein"
/standard_name= "rhPDGF-B"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 36:

GGAATTTTGT. CTTCATTAAC GGCTTTCGCT CATAAAAATG TTATGACGTT TTGCCCGCAG GCGGGAAACC ATCCACTTCA CGAGACTGAT CTCCTCTGCC

GGATCCTCA ATATGCGCCG ATACCTCAGT ATATTCGAGG TCTCTCCTT TAAGAAACGAA

AGGGCGCTTC CTTGCTAAGG ATTTCAAGG TGTTCCAAAT CTACACAAAT TGCAATAAGC

CAGTTCTAGT ATACCTCATG GCAACUGGC ACAACAAATC CAGTCGCGAT

CAACGCGAC CATTCCTCCC GTAGCTCTAC CATATTTATG TCCGCAATGAG CATCACGAGAT

ATAGCCGCAAC ACGCCTTTG GAGGGAATTA TAGAAGCTCC ATGCCGAAAT

CACCATCTG ACGCTGCGG GAATGATTT TGGGTATTG TAGTCTGTTC CAGTCAGGAA

TGGTGACGAT GAAGCTTAACT TCCACGAAAG TCCGACTCTT TCACTTCAAT TGAGGAAGAA

TACCTTCAC GTGCTTCTTT ATGCAGCTTC CGGCAACAC GATACCGATA CTCTCCATT

CGTCTTCCAT GGTTCGTTTT TTGTTGAAAG GACTAACTCA AGAATCGTCT TTCTCAAAAA

ATATCAGTT CTACCTGAGG TCTGGCAAGA AGGGCAACA CGAAGGGCAA ACAACACGAA

AAATGGGATC TCAATATTGC TGATCGCAAG AACTCTCACG AGCTCTATCT AAAAGATCCC

TTAGACGGT CTTGCCCTGG ATACACCTGT GTACGATT ATCTCGGTCT TCTTCTTTAC

CATCTGATT TTTTTTTCTA GGGATTGTG TCCATTAAAC GCCTTGCTCT CATAAATAGT

TTAGAAGCTT TGGCCTCCAG GGGGAAAC ACATCGACCA CTCTCTTCTT GCCTCTGCC

60

AGGGCGCTTC CTTGCTAAGG ATTTCAAGG TGTTCCAAAT CTACACAAAT TGCAATAAGC

CAGTTCTAGT ATACCTCATG GCAACUGGC ACAACAAATC CAGTCGCGAT

CAACGCGAC CATTCCTCCC GTAGCTCTAC CATATTTATG TCCGCAATGAG CATCACGAGAT

ATAGCCGCAAC ACGCCTTTG GAGGGAATTA TAGAAGCTCC ATGCCGAAAT

CACCATCTG ACGCTGCGG GAATGATTT TGGGTATTG TAGTCTGTTC CAGTCAGGAA

TGGTGACGAT GAAGCTTAACT TCCACGAAAG TCCGACTCTT TCACTTCAAT TGAGGAAGAA

TACCTTCAC GTGCTTCTTT ATGCAGCTTC CGGCAACAC GATACCGATA CTCTCCATT

CGTCTTCCAT GGTTCGTTTT TTGTTGAAAG GACTAACTCA AGAATCGTCT TTCTCAAAAA

ATATCAGTT CTACCTGAGG TCTGGCAAGA AGGGCAACA CGAAGGGCAA ACAACACGAA

AAATGGGATC TCAATATTGC TGATCGCAAG AACTCTCACG AGCTCTATCT AAAAGATCCC

TTAGACGGT CTTGCCCTGG ATACACCTGT GTACGATT ATCTCGGTCT TCTTCTTTAC

CATCTGATT TTTTTTTCTA GGGATTGTG TCCATTAAAC GCCTTGCTCT CATAAATAGT

TTAGAAGCTT TGGCCTCCAG GGGGAAAC ACATCGACCA CTCTCTTCTT GCCTCTGCC

120

180

240

300

360

420

480

540

600

660

720

780

840
GGAACACCG GCATCTCCAA CTTATAAGTT GGAAGAATTA AAGAAATTTCA GATTGAAGA
900
ATGAAAAAAA AAAACCGTGA AAAAAAGGST TGAACACAGT TCCGCTGAAT TATTCGCCCTA
960
CTGACTAAT AAGATATXAA AGACGATTAG TGTTGTTGAT ATACATTTAA ATCTATTCTT
1020
TAAACTTCTA AATATCTACT TTTATAGTGA GTCTTTTTTT TTTTTTAAA AAGCAAAAGA
1080
CTTGTGTTCG AATAAACAGA CATAAAACAA CACC ATG AGA TTT CCT TCA ATT
1132
Met Arg Phe Pro Ser Ile
-96 -95

TTT ACT GCA GTT TTG TTA TCC GCA GCC TCG AGC GCA TCA GCT GCA GTC
Phe Thr Ala Val Leu Phe Ala Ala Ser Ser Ala Leu Ala Ala Pro Val
-90 -85 -80 -80 -75

AAC ACT ACA GAA GAT GAA ACG GCA CAA ATT CGG GCT AAA AGA GAC
Asn Thr Thr Thr Glu Asp Glu Thr Ala Glu Ile Pro Ala Lys Arg Asp
-70 -65 -60

GCC ATT CCC GAG GAG CTC TAC GAG ATG CTT AGT GAC CAC TCG ATC CCG
Pro Ile Pro Glu Leu Leu Tyr Glu Met Leu Ser Asp His Ser Ile Arg
-55 -50 -45

TCC TTT GAT GTA TCT CAA CCC CTC CTG CAC CAC GAA GCA GCC GAA GAG GAA
Ser Phe Asp Leu Leu Gln Arg Leu Leu His Gly Asp Pro Gly Glu Leu
-40 -35 -30

GAT GGG GGC GAG TGG GAC CTG AAC ATG ACC CCC TAC CAC TCT GGA GCC
Asp Gly Ala Glu Leu Asp Leu Met Thr Arg Ser His Ser Gly Gly
-25 -20 -15

GAG CTG GAG AGC TCT TGG CGG GGC AAG AGG ACC CTG GTG CCC ACC GAG GTG
Glu Leu Ser Ser Leu Ala Arg Gly Lys Arg Ser Leu Gly Ser Leu Thr
-10 -5 1 5

ATT GCT GAG CCG ATG ATC CCG GAG TGC AAG ACG CCC ACC GAG GTG
Ile Ala Glu Pro Ala Met Ile Ala Glu Cys Thr Arg Thr Glu Val
10 15 20

TTC GAG ATC TCC CCG CCG CTC ATG GAC ACC ACC ACC GGC AAC TTC CTG
Phe Glu Ile Ser Arg Leu Ile Asp Arg Thr Asn Ala Asn Phe Leu
25 30 35 35

GTT TGG CCC CCC TCT GTG GAG GTG CAG CCC TCC GCC TCC GCC TGC AAC
Val Trp Pro Pro Cys Val Glu Val Glu Asp Ser Gly Cys Ser
40 45 50

AAC CCC AAC GTC CAG TCC CCC ACC CAG GTG CAG CTG GCA CTC TGC
Asp Arg Asn Val Glu Arg Pro Thr Glu Val Glu Leu Arg Pro Val
55 60 65 70

CAG GTG AAG ATC GAG ATT GTG CGG AAG AAG CCA ATC TTT AAG AAG
Gln Val Arg Ile Glu Ile Val Arg Lys Pro Ile Arg Lys Lys
75 80 85

GCC ACC GTG AGC CTG GAA GAC CAC CTG GCC TGC AAG TGT GAG ACA GTG
Ala Thr Val Thr Leu Gly Asp His Leu Ala Cys Glu Thr Val
90 95 100

GCA GCT GCA CGG CCT GTG ACC TAA TAG CATGCTCAGC TTTGTTCCCA
1023
Ala Ala Ala Arg Pro Val Thr
105 110

CTGACCTTT AAGCTCATCA AAAAAAAATA TACCTTTTAT TTTGCTGAAAA AAAAAATTTT
1084
TTCCGACTAA ATACTCCTTT CTATTCTTAC GACCTTACG ATACTTAAAA
1142
TAGCCACCC ATATCGATTC AAAAAACT AAAAAATTTT GCACTTTGCC
1200
TTCAGATTTG TTATATAGCC ATATAATTTA TTTTTTTTAG AGCTAAAAAAAA AGATGTAAGTG
1258
GAATGCAACT CTAAAAGAACT TGCGATCC
1316
(2) INFORMATION FOR SEQ ID NO: 37:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 285 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 37:

Met Arg Phe Pro Ser Ile Phe Thr Ala Val Leu Phe Ala Ala Ser Ser
-96  -95  -90  -85
Ala Leu Ala Ala Pro Val Asn Thr Thr Glu Asp Glu Thr Ala Gln
-80  -75  -70  -65
Ile Pro Ala Lys Arg Asp Pro Ile Pro Glu Glu Leu Tyr Glu Met Leu
-60  -55  -50
Ser Asp His Ser Ile Arg Ser Phe Asp Asp Leu Gln Arg Leu Leu His
-45  -40
Gly Asp Pro Gly Glu Ala Asp Gly Ala Leu Asp Leu Asn Met Thr
-30  -25  -20
Arg Ser His Ser Gly Gly Leu Glu Ser Leu Ala Arg Gly Lys Arg
-15  -10  -5
Ser Leu Gly Ser Leu Thr Ile Ala Gla Pro Ala Met Ile Ala Glu Cys
1    5   10  15
Lys Thr Arg Thr Glu Val Phe Glu Ile Ser Arg Arg Leu Ile Asp Arg
20   25   30
Thr Asn Ala Asn Phe Leu Val Trp Pro Pro Cys Val Glu Val Gln Arg
35  40   45
Cys Ser Gly Cys Cys Asn Asn Arg Asn Val Gln Cys Arg Pro Thr Gln
50  55   60
Val Gln Leu Arg Pro Val Gln Val Arg Lys Ile Glu Ile Val Arg Lys
65  70   75  80
Lys Pro Ile Phe Lys Lys Ala Thr Val Thr Leu Glu Asp His Leu Ala
85  90  95
Cys Lys Cys Glu Thr Val Ala Ala Ala Arg Pro Val Thr
100 105

(2) INFORMATION FOR SEQ ID NO: 38:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 480 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid
(A) DESCRIPTION: /desc = "chimeric DNA molecule"

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens/Saccharomyces cerevisiae

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 1..471

(ix) FEATURE:
(A) NAME/KEY: misc.feature
(B) LOCATION: 1..255
(D) OTHER INFORMATION: /function= "mediates protein secretion" /
/product= "Yeast alpha factor leader peptide" /
/standard_name= "Alpha factor signal/leader sequence"

(ix) FEATURE:
(A) NAME/KEY: mat_peptide
(B) LOCATION: 256..471
(D) OTHER INFORMATION: /product= "rhIGF-I-A protein" /
/standard_name= "rhIGF-I-A"
ATG Met -85 GCA Ala
ATT Ile
GAT Asp
TTT Phe
CAG Glin
GAT Asp
CCA Pro
GTT Wall
TAC Tyr 60

ATT GCA GCT GTC ATC GAT TTA GAA GGG GAT TCC
Ile Pro Ala Glu Ala Val Ile Gly Tyr Leu Asp Leu Gly Asp Phe
-50 -45 -40

GAT GCT GCT GGT CCA TTT TCC AAC AGC ACA AAT AGG TTA TGT
Asp Val Ala Val Leu Pro Phe Ser Asn Ser Thr Asn Asp Gly Leu Leu
-35 -30 -25

TAT AAT ACT ACT ATT GCC ATT GCT AAT AAA GAA GGA GTA
Phe Ile Asn Thr Thr Thr Ala Ala Ala Asp Leu Gly Leu Gly Val
-20 -15 -10

CAS CTC GAT AAA AGA GCT CCA GAA ACC TTG TGT GCT GAA TGG TCC
Gln Leu Asp Asp Arg Gly Pro Glu Thr Leu Cys Gly Ala Leu Leu Val
-5 1 5 10

GAT GCT TGG CAA TTC GAT GAC AGA GGT TAC TAC AAC AAG
Asp Ala Leu Gln Phe Val Cys Gly Asp Arg Gly Phe Tyr Phe Asn Lys
15 20 25

CCA ACC GGT TAC GGT TCT TCT TCT AAG AAG GCT CCA CAA ACC GGT ATC
Pro Thr Gly Tyr Gly Ser Ser Arg Ala Pro Gin Thr Gly Ile
30 35 40

GTT GAC GAA TGT TGT TCT AAG TGT TGT GAC TTG AGA TGG GTA
Val Asp Gly Cys Cys Phe Arg Ser Cys Asp Leu Arg Leu Gly Met
45 50 55

TAC TGT GCT CCA TTG AAG CCA GCT GGT TCG ACT TTA TGT GCG TTT
Tyr Cys Ala Pro Leu Lys Pro Ala Lys Ser Ala * *
60 65 70

(2) INFORMATION FOR SEQ ID NO: 39:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 155 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 39:
Met Arg Phe Pro Ser Ile Phe Thr Ala Val Leu Phe Ala Ala Ser Ser
-85 -80 -75 -70

Ala Leu Ala Ala Pro Val Asn Thr Thr Glu Asp Glu Thr Ala Gin
-65 -60 -55

Ile Pro Ala Glu Ala Val Ile Gly Tyr Leu Asp Leu Gly Asp Phe
-50 -45 -40

Asp Val Ala Val Leu Pro Phe Ser Asn Ser Thr Asn Asp Gly Leu Leu
-35 -30 -25

Phe Ile Asn Thr Thr Ala Ser Ile Ala Lys Gly Glu Gly Val
-20 -15 -10

Gln Leu Asp Lys Arg Gly Pro Glu Thr Leu Cys Gly Ala Leu Leu Val
-5 1 5 10

Asp Ala Leu Gin Phe Val Cys Gly Asp Arg Gly Phe Tyr Phe Asn Lys
15 20 25

Pro Thr Gly Tyr Gly Ser Ser Arg Arg Ala Pro Gin Thr Gly Ile
30 35 40
Val Asp Glu Cys Cys Phe Arg Ser Cys Asp Leu Arg Arg Leu Glu Met
45 50 55

Tyr Cys Ala Pro Leu Lys Pro Ala Lys Ser Ala
60 65 70

(2) INFORMATION FOR SEQ ID NO: 40:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 621 base pairs
(B) TYPE: nucleic acid
(C) STRANDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid
(A) DESCRIPTION: /desc = "Chimeric DNA molecule"

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens/Saccharomyces cerevisiae

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 1..579

(ix) FEATURE:
(A) NAME/KEY: misc_feature
(B) LOCATION: 256..471
(D) OTHER INFORMATION: /function= "mediates secretion of protein"
/product= "3'end of yeast alpha factor leader peptide"
/standard_name= "alpha factor leader/signal sequence"

(ix) FEATURE:
(A) NAME/KEY: mat_peptide
(B) LOCATION: 256..471
(D) OTHER INFORMATION: /product= "rhIGF-I-A protein"
/standard_name= "rhIGF-I-A"

(ix) FEATURE:
(A) NAME/KEY: transit_peptide
(B) LOCATION: 472..579
(D) OTHER INFORMATION: /function= "mediates protein transport/translation"
/product= "IGF-I-A propeptide"
/standard_name= "IGF-I-A prosequence"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 40:

ATG AGA Met Arg -85 GCA TTA Ala Leu
ATT CCG Ile Pro 471
GAT GTT Asp Wall 426
TTT ATA Phe Ile 381
GAG CTT Val Ile
CAC GGT Met 346
AGA Arg 311
TAC Tyr -266
GGT GAA Glu 281
AGA Arg 246
TAC Tyr -211
GGT GAA Glu 186
AGA Arg 151
TAC Tyr -116
GGT GAA Glu 91
AGA Arg 56
TAC Tyr -21
GGT GAA Glu 4
AGA Arg 4

(continued)
What is claimed is:
1. A vector comprising a nucleotide sequence that comprises in the 5' to 3' direction and operably linked (a) a yeast-recognized transcription and translation initiation region, (b) a coding sequence for a hybrid precursor polypeptide, and (c) a yeast-recognized transcription and translation termination region, wherein said hybrid precursor polypeptide comprises:

\[ \text{5'-SP-(PS)}_{\text{H}}\text{-(LP-PS)}_{\text{R}}\text{-}(\text{NPRO}_{\text{H}}\text{-PS})_{\text{M}}\text{-}(\text{PS-CPRO}_{\text{R}})\text{ HP}_{\text{X}}\text{-3'}}\]
SP comprises a signal peptide sequence for a yeast secreted protein;

PS comprises a preferred processing site cleaved in vivo by a yeast proteolytic enzyme;

LP comprises a leader peptide sequence for a yeast secreted protein;

NPRO<sub>MHP</sub> comprises a native N-terminal propeptide sequence of a mature heterologous mammalian protein of interest;

MHP comprises a peptide sequence for said mature heterologous mammalian protein of interest;

CPRO<sub>MHP</sub> comprises a native C-terminal propeptide sequence of said mature heterologous mammalian protein of interest; and

n-1, n-2, n-3, and n-4 independently = 0 or 1;

wherein said mammalian protein is a native PDGF-BB protein or variant thereof, wherein said variant has an amino acid sequence that has at least 70% sequence identity to the amino acid sequence of said native PDGF-BB protein, wherein said processing sites allow for proteolytic processing of said precursor polypeptide to said mature protein in vivo by a yeast host cell from the genus Kluyveromyces, and wherein at least n-3 or n-4 > 1.

2. The vector of claim 1, wherein n-2=1, n-3=1, and n-4=0.

3. The vector of claim 1, wherein LP is a truncated leader peptide sequence.

4. The vector of claim 1, wherein said yeast host cell is 

<i>Kluyveromyces lactis</i>.

5. A yeast host cell stably transformed with the vector of claim 1, wherein said yeast cell is from the genus Kluyveromyces.

6. A yeast host cell stably transformed with the vector of claim 2, wherein said yeast cell is from the genus Kluyveromyces.

7. A yeast host cell stably transformed with the vector of claim 3, wherein said yeast cell is from the genus Kluyveromyces.

8. A yeast host cell stably transformed with a vector, said vector comprising a nucleotide sequence that comprises in the 5' to 3' direction and operably linked (a) a yeast-recognized transcription and translation initiation region, (b) a coding sequence for a hybrid precursor polypeptide, and (c) a yeast-recognized transcription and translation termination region, wherein said hybrid precursor polypeptide has an amino acid sequence selected from the group consisting of the amino acid sequence set forth in SEQ ID NO: 27 and an amino acid sequence having at least 70% sequence identity to the amino acid sequence set forth in SEQ ID NO: 27, wherein said yeast cell is from the genus Kluyveromyces.

9. The yeast host cell of claim 8, wherein said yeast host cell is 

<i>Kluyveromyces lactis</i>.

10. A method for expression of a heterologous protein and its secretion in the biologically active mature form using a yeast host cell from the genus Kluyveromyces as the expression system, said method comprising transforming said yeast cell with a vector comprising a nucleotide sequence that comprises in the 5' to 3' direction and operably linked

(a) a yeast-recognized transcription and translation initiation region, (b) a coding sequence for a hybrid precursor polypeptide, and (c) a yeast-recognized transcription and translation termination region, wherein said hybrid precursor polypeptide comprises:

5'-SP-(PS)<sub>n-1</sub>-(LP-PS)<sub>n-2</sub>-(NPRO<sub>MHP</sub>-PS)<sub>n-3</sub>-MHP-(PS-CPRO<sub>MHP</sub>-

wherein:

SP comprises a signal peptide sequence for a yeast secreted protein;

PS comprises a preferred processing site cleaved in vivo by a yeast proteolytic enzyme;

LP comprises a leader peptide sequence for a yeast secreted protein;

NPRO<sub>MHP</sub> comprises a native N-terminal propeptide sequence of a mature heterologous mammalian protein of interest;

MHP comprises a peptide sequence for said mature heterologous mammalian protein of interest;

CPRO<sub>MHP</sub> comprises a native C-terminal propeptide sequence of said mature heterologous mammalian protein of interest; and

n-1, n-2, n-3, and n-4 independently = 0 or 1;

wherein said mammalian protein is a native PDGF-BB protein or variant thereof, wherein said variant has an amino acid sequence that has at least 70% sequence identity to the amino acid sequence of said native PDGF-BB protein, wherein said processing sites allow for proteolytic processing of said precursor polypeptide to said mature protein in vivo by said yeast host cell, and wherein at least n-3 or n-4 > 1; and cultivating said transformed cell such that said protein is expressed.

11. The method of claim 10, wherein said processing sites are dipeptides.

12. The method of claim 11, wherein said dipeptides are

5'-Lys-Arg-3'.

13. The method of claim 10, wherein said yeast host cell is 

<i>Kluyveromyces lactis</i>.

14. A method for expression of a heterologous protein and its secretion in the biologically active mature form using a yeast host cell from the genus Kluyveromyces as the expression system, said method comprising transforming said yeast cell with a vector comprising a nucleotide sequence that comprises in the 5' to 3' direction and operably linked

(a) a yeast-recognized transcription and translation initiation region, (b) a coding sequence for a hybrid precursor polypeptide, and (c) a yeast-recognized transcription and translation termination region, wherein said hybrid precursor polypeptide comprises an amino acid sequence selected from the group consisting of the amino acid sequence set forth in SEQ ID NO: 27 and an amino acid sequence having at least about 70% sequence identity to the amino acid sequence set forth in SEQ ID NO: 27; and cultivating said transformed cell such that said protein is expressed.

15. The method of claim 14, wherein said yeast host cell is 

<i>Kluyveromyces lactis</i>.